



## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification <sup>6</sup> : <b>C12N 15/86, 5/10, 15/67</b>		A1	(11) International Publication Number: <b>WO 97/08330</b>
			(43) International Publication Date: 6 March 1997 (06.03.97)
(21) International Application Number: PCT/GB96/02061 (22) International Filing Date: 23 August 1996 (23.08.96) (30) Priority Data: 9517263.1 23 August 1995 (23.08.95) GB (71) Applicant (for all designated States except US): CANCER RESEARCH CAMPAIGN TECHNOLOGY LIMITED [GB/GB]; Cambridge House, 6-10 Cambridge Terrace, Regent's Park, London NW1 4JL (GB). (72) Inventors; and (75) Inventors/Applicants (for US only): COLLINS, Mary, Katharine, Levinge [GB/GB]; Flat 4, Philips House, 52 Goodge Street, London W1P 1FP (GB). WEISS, Robin, Anthony [GB/GB]; 25 Cyprus Avenue, London N3 1SS (GB). TAKEUCHI, Yasuhiro [JP/GB]; 141 Elborough Street, London SW18 5DS (GB). COSSET, François-Lois [FR/FR]; 32, rue L.-Thévenet, F-69004 Lyon (FR). (74) Agents: CALDERBANK, T., R. et al.; Mewburn Ellis, York House, 23 Kingsway, London WC2B 6HP (GB).		(81) Designated States: AL, AM, AT, AU, AZ, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, US, UZ, VN, ARIPO patent (KE, LS, MW, SD, SZ, UG), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG). Published With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.	
(54) Title: EXPRESSION SYSTEMS			
<p>pol gene...TCT AGA CTG ACA TG G CGC          GTT CAA CGC TCT CAA AAC CCC TTA          AAA ATA AGG TTA ACC CGC GAG GCC          CCC TAA          tcccttaattcttctgaigctcagaggggtcagtac          tgcctgcgcgggtccagtgccgcccagccgcccacc          ATG AAA ACA TTT AAC ATT TCT... bsr gene</p>			
Schematic structure of CeB expression vector			
(57) Abstract			
<p>The invention relates to new expression systems and in particular to an expression system in which a gene of interest is expressed at an optimal level. The invention provides a recombinant expression vector comprising a gene of interest and a selectable marker gene, wherein the selectable marker gene is arranged downstream of the gene of interest and a stop codon associated with the gene of interest is spaced from a start codon of said selectable marker gene at a distance which is sufficient to ensure that translation reinitiation is required before said selectable marker protein is expressed from the corresponding MRNA. Examples of such expression systems are vector viral packaging cell lines and a number of preferred cell lines have been identified.</p>			

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	GB	United Kingdom	MW	Malawi
AT	Austria	GE	Georgia	MX	Mexico
AU	Australia	GN	Guinea	NE	Niger
BB	Barbados	GR	Greece	NL	Netherlands
BE	Belgium	HU	Hungary	NO	Norway
BF	Burkina Faso	IE	Ireland	NZ	New Zealand
BG	Bulgaria	IT	Italy	PL	Poland
BJ	Benin	JP	Japan	PT	Portugal
BR	Brazil	KE	Kenya	RO	Romania
BY	Belarus	KG	Kyrgyzstan	RU	Russian Federation
CA	Canada	KP	Democratic People's Republic of Korea	SD	Sudan
CF	Central African Republic	KR	Republic of Korea	SE	Sweden
CG	Congo	KZ	Kazakhstan	SG	Singapore
CH	Switzerland	LI	Liechtenstein	SI	Slovenia
CI	Côte d'Ivoire	LK	Sri Lanka	SK	Slovakia
CM	Cameroon	LR	Liberia	SN	Senegal
CN	China	LT	Lithuania	SZ	Swaziland
CS	Czechoslovakia	LU	Luxembourg	TD	Chad
CZ	Czech Republic	LV	Latvia	TG	Togo
DE	Germany	MC	Monaco	TJ	Tajikistan
DK	Denmark	MD	Republic of Moldova	TT	Trinidad and Tobago
EE	Estonia	MG	Madagascar	UA	Ukraine
ES	Spain	ML	Mali	UG	Uganda
FI	Finland	MN	Mongolia	US	United States of America
FR	France	MR	Mauritania	UZ	Uzbekistan
GA	Gabon			VN	Viet Nam

## Expression systems

The present invention relates to new expressions systems,  
and in particular to expression systems in which a gene of  
interest is expressed at an optimal level. Particular  
examples of such expression systems are retroviral packaging  
cell lines and a number of preferred cell lines have been  
identified.

The ability of eukaryotic and prokaryotic ribosomes to  
reinitiate translation at an internal start codon within an  
mRNA sequence has previously been recognised. Studies have  
been reported in which the efficiency of the process, which  
is generally regarded as being low, has been connected with  
the length of the intercistronic sequence (Kozak (1987)  
Mol. Cell Biol. 7, 3438-3445). Selection of this sequence  
or spacer as 70bp in length, and containing no other start  
codons, has been previously reported as being optimal for  
reinitiation in a eukaryotic cell line (Cosset F-L.,  
Virology (1991) 185, 862).

The applicants have found a way in which the inefficiency  
associated with the translation reinitiation process can be  
used to good effect.

According to the present invention there is provided a  
recombinant expression vector comprising a gene of interest  
and a selectable marker gene, wherein the selectable marker  
gene is arranged downstream of the gene of interest and a  
stop codon associated with the gene of interest is spaced  
from a start codon of said selectable marker gene at a  
distance which is sufficient to ensure that translation re-  
initiation is required before said selectable marker protein  
is expressed from the corresponding mRNA.

The invention further provides a process for producing cell lines in which a gene of interest is expressed, which process comprises transforming host cells with an expression vector comprising said gene of interest and a selectable marker gene, wherein the selectable marker gene is arranged downstream of the gene of interest and a stop codon associated with the gene of interest is spaced from a start codon of said selectable marker gene at a distance which is sufficient to ensure that translation re-initiation is required before said selectable marker protein is expressed from the corresponding mRNA, and selecting those cells where expression of the selectable marker gene may be detected.

Since re-initiation of translation is a relatively inefficient process, this means that the selectable marker protein will be expressed at lower levels than the product of the gene of interest. When the marker protein is expressed at detectable levels, the gene of interest will be expressed at higher levels. This will ensure that during the subsequent selection procedure, only those cell clones which express the gene of interest at higher or optimal levels will survive. Low expressing clones will be eliminated by the selection process.

Cells transformed with the above-described expression vectors form a further aspect of the invention.

The host cells are suitably eukaryotic or prokaryotic host cells, preferably eukaryotic host cells.

The number of nucleotides in the space between the stop codon of the gene of interest and the start codon of the selectable marker will suitably be in the range of from 20-200 nucleotides, preferably from 60-80 nucleotides, even more preferably 70-80 nucleotides.

The vectors used in the process of the invention may be any of the known types, for example expression plasmids or viral vectors.

5 Selected cells may be cultured and if required, the protein product of the gene of interest isolated from the culture using conventional techniques. Alternatively, expression of the gene of interest may result in other desired effects, for example, where the gene of interest is included as part  
10 of a viral packaging construct.

Some experimental and clinical gene transfer protocols require the design of gene transfer vectors suitable for *in vivo* gene delivery (Miller, A.D. 1992. Nature 357:455-  
15 460). Retroviral vectors are attractive candidates for such applications, because they can provide stable gene transfer and expression (Samarut J. et al., Meth. Enzymol. in press) and because packaging cells have been designed which produce non-replication competent viruses (Miller A.D (1990) Hum Gene Ther. 1 5-14). However currently available  
20 recombinant retroviruses suffer from a number of drawbacks.

Packaging cell lines provide in trans the retroviral proteins encoded by the gag, pol, and env genes required to  
25 obtain infectious retroviral particles. The gag and pol products are respectively the structural components of the virion cores and the replication machinery (enzymes) of the retroviral particles whereas the env products are envelope proteins responsible for the host-range of the virions and  
30 for the initiation of infection and for sensitivity to humoral factors. An ideal packaging cell line should produce retroviruses that only contain the retroviral vector genome, and absolutely no replication-competent genomes or defective genomes encoding some of the viral structural  
35 genes.

A number of packaging cell lines designed for human gene transfer have been designed in the past by introducing plasmid DNAs which contain "helper genomes" encoding gag, pol and/or env genes into cells.

5

Retroviral packaging cell lines are cells that have been engineered to provide in trans all the functions required to express infectious retroviral vectors. A helper genome (or construct or unit), is herein also referred to as "retroviral packaging construct (or unit)" or "packaging-deficient construct (or genome unit)" or "gag-pol/env expression plasmids".

10

Much efforts has been made to design strategies to optimize the helper-genomes in order (i) to get the highest production of retroviral packaging functions (which correlates with infection titers of retroviral particles) and (ii) to minimise the chance that the helper genome can be transmitted via the viral particles (which may lead to emergence of unwanted retroviral forms).

15

20

The first of these packaging cell lines used full length retroviral genomes as helper genomes that had been crippled for important cis-regulated replicative functions (reviewed in Miller, Hum. Gene. Ther. 1:5-14 1990). In order to reduce the possibility of occurrence of replication-competent viruses and of transfer of virus structural genes, a second generation of safer packaging cell lines has been designed by using two separate and complementary helper genomes which express either gag-pol or env and are packaging-deficient (Miller supra).

25

30

The cells into which these helper genomes were introduced were isolated by cotransfecting them with plasmids encoding selectable markers. However, as no selection was applied on

35

the packaging-deficient retroviral genome itself, the helper functions can be lost during the passages of the cells in culture and the current packaging systems provide limited titers of infectious retroviral vectors, usually only of the order of  $10^5$ - $10^6$  infectious units i.u./ml. Indeed the cotransfection with a plasmid encoding a selectable marker does not directly select the best gag-pol-env-expressing cells.

The invention further provides a retroviral packaging cell line comprising a host cell transformed with (i) a packaging deficient construct which expresses a viral gag-pol gene and a first selectable marker gene, and/or (ii) a packaging-deficient construct which expresses a viral env gene and a second selectable marker gene; wherein a start codon of the first and second selectable markers are spaced from the stop codons of the viral gag-pol gene and the viral env gene respectively by a distance which ensures that reinitiation of mRNA translation is required for expression of marker protein product of said first and/or second selectable marker gene.

The retroviral packaging cell line may be obtained by the above described process which will involve selecting transfected cells which express said first and/or second marker genes.

By using helper constructs which are directly selectable and which provide for high expression of the viral gene, high titre retroviral vectors may be obtained.

Helper constructs for use in the process form a further aspect of the invention.

The retroviral vectors prepared from the conventional

packaging cell lines are usually not contaminated by replication-competent retroviruses (RCRs). However, recombinant amphotropic murine retroviruses have been shown to arise spontaneously from certain packaging cell lines.

5 The generation of such RCRs involves recombination at least between gag-pol/env packaging sequence and vector sequences (Cosset et al., Virology, (1993) 193:385-395).

Recombinant RCRs have been associated with the development of lymphomas in some severely immunosuppressed monkeys (Donahue et al., J. Exp Med (1992) 176: 1125-1135). In addition, retroviral vector preparations may also contain, at low frequencies, retroviruses coding for functional envelope glycoproteins (Kozak and Kabat, 1990, J. Virol. 64: 3500-3508) or for gag-pol proteins. Although the

10 pathogenicity of these gag-pol or env recombinant retroviruses is probably low, more evolved recombinant retroviruses with higher pathogenic potential may occur when injected in vivo, by recombination and/or complementation of

15 the initial recombinant viruses with some endogenous retroviruses.

20

In a preferred embodiment of the retroviral packaging cell lines of the invention, the overlapping sequences between the genomes of the retroviral vector and the helper

25 construct are reduced, for example as compared to constructs such as CRIPenv and CRIPAMgag (Danos et al., Proc. Natl. Acad. Sci USA 85: 6460-6464). In particular, the viral sequences in the helper construct are reduced, for example,

30 not only the packaging sequence but also the 3' Long Terminal Repeat (LTR), the 3' non-coding sequence and/or the 5' LTR may be eliminated.

The possibility of generation of such RCRs and recombinant retroviruses can be reduced by reducing the overlapping

35



sequences between the genomes of both the retroviral vector and the helper construct.

5 Conventional retroviral vectors are strongly inactivated by human serum which makes them of limited or no use for in situ gene transfer in gene therapy applications. It has previously been shown that inactivation by complement in human serum is controlled by the cell line used to produce the virions and by viral envelope determinants (Takeuchi et al., J. Virol (1994) 68:8001-8007). In particular, 10 inactivation is caused by some properties of the cell lines that have been used to construct the packaging cells (NIH-3T3) and also by viral determinants located in the retroviral envelope as shown (Takeuchi et al., J. Virol (1994) 68:8001-8007). In vivo gene delivery is an important 15 goal for a number of human gene therapy strategies.

The applicants have found that certain cell lines form preferred packaging cell lines. 20

Particularly preferred packaging cell lines are the HT1080 line, the TE671 line, the 3T3 line, the 293 line and the Mv-1-Lu line. One example of retroviral packaging cells that will produce complement-resistant virus comprise human 25 HT1080 cells and express RD114 envelope. Such cells form a preferred aspect of the invention.

Packaging cell lines according to the invention provide 50-100 fold increased titers of retroviral vectors as compared to conventional packaging cell lines. Retroviral vectors 30 provided by these new cells are safe, in terms of generation of RCRs, and considerably more resistant to inactivation by human complement.

35 Packaging cell lines according to the invention may be able

to transduce helper-free, human complement-resistant retroviral vectors at titers consistently higher than  $10^7$  i.u./ml.

5 Suitable semi-packaging cell lines in accordance with the invention are those which express only the gag-pol genes. Such cell lines may suitably be derived from TE671, MINK Mv-1-Lu, HT1080, 293 or NIH-3T3 cells by introduction of plasmid CeB (the MoMLV gag-pol expression unit).

10

Particularly preferred expression vectors in accordance with the invention for use in retroviral packaging cell lines are those which include MLV gag and pol genes such as CeB. Other plasmids may include gag and pol genes from other  
15 retroviruses or chimeric or mutated gag and pol genes.

Various viral and retroviral envelope genes may be included in the plasmids such as MLV-A envelope, GALV envelope, VSV-G protein, BaEV envelope, RD114 envelope and chimeric or  
20 mutated envelopes. Plasmids which include the RD114 env gene such as FBdelPRDSAF as illustrated hereinafter, provide one example of suitable constructs.

The novel retroviral packaging cells described hereinafter,  
25 have been designated FLY cells, and may be designed for in vivo gene delivery.

Considerable variations were found between the various cell lines screened for their ability to release type C mammalian retroviruses. In addition, few cell lines were able to  
30 produce retroviruses completely resistant to human complement. Based on these two criteria, human fibrosarcoma HT1080 and rhabdomyosarcoma TE671 cells were selected for optimum construction of packaging cells.

35

Other studies have shown the importance of endogenous retrovirus expression in the generation of recombinant retroviruses from retroviral packaging lines (Ronfort et al., Virology, (1995), 207, 271-275, Vanin, E.F. et al., J Virol (1994) 68:4241-4250.). The co-packaging of an endogenous genome and a vector can lead to emergence of recombinant retroviruses (Vanin et al., supra). Recombination involves template switching during reverse transcription of such hybrid retroviruses (Hu et al., Science, (1990) 250:1227) and homologies between the two genomes considerably enhance the frequency of reverse transcriptase jumps (Zhang et al., J. Virol. (1994) 68: 2409-2414). Therefore an ideal packaging cell line should not express endogenous MLV-like (or type C retrovirus-like) retroviral genomes which can be packaged by type C gag proteins (Scadden et al., J. Virol. (1990) 64: 424-427, Torrent et al., J. Mol. Biol. (1994) 240 434-444).

Packaging of human endogenous retroviral RNA was not detected in TELCeB and FLY packaging cells when virion associated RNA was analysed by RT-PCR using generic primers. HT1080- and TE671 derived packaging cell lines may be safer in this respect than those generated from NIH3T3 cells, such as GP+EAM12 cells, which are known to express and package sequences related to type C retroviruses (Scadden et al. supra).

To generate the FLY packaging cell lines, HT1080 cells were transfected with gag-pol and env expression plasmids designed to optimise viral protein expression. Direct selection for viral gene expression was achieved in accordance with the invention by expression of a selectable marker gene by re-initiation of translation of the mRNA expressing the viral proteins. This strategy resulted in packaging cell lines capable of producing extremely high

titer viruses. Furthermore, long-term expression of packaging functions can be maintained in these cells. Many unnecessary viral sequences were eliminated from the packaging constructs to reduce the risk of helper virus generation; indeed the final packaging cells did not produce helper virus, in that no replication competent virus (RCR) could be detected per  $10^7$  vector particles.

The FLY packaging cells described herein are safer than, for example, psiCRIP cells, at least for generation of env recombinant retroviruses as is illustrated in Table 4 hereinafter, probably because less retroviral sequences overlapping with the vector were present in the present env-expression plasmid. Few reports have addressed the question of the characterization of recombinant retroviruses (RVs) (Cosset, F.L., et al., Virology (1993) 193:385-395). It is possible that such RVs could not be detected in previous packaging cell lines due to lower overall titers. RVs are defective in normal cell culture conditions but are likely to evolve to replication competent viruses if they are allowed to replicate in cells complementing their expression like co-cultivated packaging cells (Bestwick et al., Proc. Natl Acad Sci USA, (1988) 85: 5404-5408, Cosset et al., (1993) supra).

In preferred retroviral packaging systems according to the invention, RVs are eradicated for example by removal of viral LTRs from the packaging construct.

Consistent with our previous studies (Takeuchi, Y., et al., J Virol (1994) 68:8001-8007), LacZ(RD114) and lacZ(MLV-A) pseudotypes produced from HT1080 and TE671 cells were more resistant to human complement than LacZ(RD114) or LacZ(MLV-A) pseudotypes produced by 3T3 or dog cells. It was therefore decided to use RD114 and MLV-A env genes to

generate recombinant virions with MoMLV cores.

5 The sequence of RD114 env gene was determined and is shown  
in Figure 4. It was found to be very close to BaEV (baboon  
endogenous virus) a type C retrovirus (Benveniste, R.E. et  
al., Proc. Natl. Acad. Sci. USA (1973) 70:3316-3320; Kato,  
S. et al., Japan. J. Genet. (1987) 62:127-137) with an  
10 envelope gene displaying similarities to the external part  
of type D simian retroviruses (SRVs). RD114 uses the SRV  
receptor on human cells (Sommerfelt & Weiss, Virology  
(1990) 176:58-69; Sommerfelt, M.A. et al., J Virol (1990)  
64:6214-6220) making the FLY packaging cells with RD114  
envelope capable of generating virions with different  
15 tropism. Retroviral vectors prepared so far for human gene  
therapy have used either MLV-A or GALV (gibbon ape leukemia  
virus) envelopes which display some similarities (Battini,  
J.L., et al., J Virol. (1992) 66:1468-1475) and which use two  
related cell surface receptors for infection (Miller, D.G.  
et al., J Virol (1994) 68:8270-8276). Differences in tissue-  
20 specific expression of MLV-A or GALV receptors have been  
reported (Kavanaugh et al., Proc Natl Acad Sci USA (1994)  
91:7071-7075).

25 The invention will now be particularly described by way of  
example with reference to the accompanying drawings in  
which:

Figure 1. illustrates the structure and expression of CeB.  
The env gene (Xbal-Clal) of plasmid pCRIP was removed and  
30 was replaced by coininsertion of the two fragments Xbal-Sfil  
(restriction sites underlined) from pOXEnv and a Sfil-Clal  
PCR product containing the bsr selectable marker. This  
results in positioning the bsr start codon (shadowed) 74 bp  
downstream to the pol stop codon (bold).

35

Open triangle are start codons (gag and bsr), black triangles are stop codons (pol and bsr). The shadowed triangle is the start codon of env, in the same reading frame with that of bsr. SD and SA are the splice donor and splice acceptor sites.

Figure 2 illustrates the structure and expression of FbdelpASAF.

Immediately after the stop codon of env (bold) was inserted a non retroviral KasI-NcoI (restriction sites underlined) linker which positions the phleo start codon (shadowed) 76 bp downstream.

Open triangle are start codons (env and phleo), black triangles are stop codons (env and phleo). SD and SA are the splice donor and splice acceptor sites.

Figure 3 illustrates plasmids for expression of Amphi, Eco, RD114, Xeno, 10A1, GALV, VSV-G and FeLV envelopes.

All genes are expressed in the same backbone as detailed in fig. 2. The BglII sites for ecotropic (MoMLV strain), 10A1, xenotropic (NZB.1.V6 strain) and amphotropic (4070A strain), the NdeI site of RD114 (SC3C strain, the BamHI site for both FeLV and GALV were used as 5' ends, and linked to MscI site immediately after the splice donor site in the leader of FB29 LTR.

Figure 4 shows the sequence of the RD114 env gene (SEQ ID No 1).

Figure 5 shows the genetic structure of gag-pol constructs. Initiation (◊) and termination (▼) codons are shown. The thick dotted line below each construct shows MLV-derived sequences. Nucleotide positions of MLV-derived sequences are shown according to: Shinnick et al. (1981) (from nt 1 to nt 6000 with deletion of the packaging signal (DY) from Bali

(nt 215) to PstI (nt 568), and with some further MoMLV sequences in both CeB and CeB DS- from nt 7676 to nt 7938. gag-pol and bsr genes were expressed from the same transcription unit using the either a retroviral promoter (Mo LTR) or a non retroviral promoter (hCMV) and non retroviral polyadenylation sequence (polyA). Splice donor (SD) and acceptor (SA) sites are indicated. The thin line denotes retroviral non coding sequences. The thick line shows the rabbit beta-1 globin intron B. The position of some restriction sites is indicated.

The nucleic acid sequences of portions of constructs (as shown in Figure 5 (boxed areas)) are displayed for CeB (SEQ ID No 2, Figure 6), hCMV+intron (SEQ ID No 3, Figure 7) and hCMV+intronkaSD (SEQ ID No 4, Figure 8).

The nucleic acid sequences of portions of constructs (as shown in Figure 3 (boxed areas)) are displayed for FbdeIPASAF (SEQ ID No 5, Figure 9), FbdeIPMOSAF (SEQ ID No 6, Figure 10), FbdeIPGASAF (SEQ ID No 7, Figure 11), FbdeIPRDSAF (SEQ ID No 8, Figure 12) and CMV10A1 (SEQ ID No 9, Figure 13) are shown.

The components of the viral particles are produced by two independent expression plasmids (gag-pol or env) which also contain selectable markers (bsr or phleo) expressed from the same transcriptional units as gag-pol or env (figs. 1 & 2). The selectable markers are located downstream to gag-pol or env genes and there is an optimal distance between the stop codon of the upstream reading frames and the start codon of the selectable genes that should allow re-initiation of translation (Kozak, Mol Cell Biol. (1987) 7, :3438-3445). Because there is no "Kozak" sequence (Kozak, Cell, (1986) 44: 283-292) required for a normal initiation of translation for

the marker gene, they can only be expressed by re-initiation of translation after the upstream viral gene has been successfully expressed. Consequently and also because re-initiation of translation is a poorly efficient process,  
5 after transfection of these plasmids, cells resistant to the drugs corresponding to those selectable genes express high levels of the viral proteins.

10 To avoid viral transmission of these "helper" genomes the constructs used suitably have the classical deletions of both the packaging sequence located in the leader region and of the 3'LTR, the latter being replaced by SV40 polyadenylation sequences (Figs 1 & 2).

15 Plasmid CeB is the MoMLV gag-pol-expression unit. It derives from pCRIP, a plasmid used to generate the constructs introduced in the CRIP and CRE packaging cell lines (Danos and Mulligan, 1988). As shown in fig. 1 for generation of plasmid CeB the env gene of pCRIP has been  
20 deleted mostly and the bsr selectable marker, -encoding a protein conferring resistance to blasticidin (Izumi et al., Experimental Cell Research (1991) 197, 229-233)- has been inserted downstream to pol gene. There are exactly 74 bp with no ATG triplets between the stop codon of pol and the  
25 start codon of bsr, this allows its expression by re-initiation of translation on the gag-pol mRNA, after translation of the gag-pol reading frame.

30 FbdelpASAF is a plasmid expressing the amphotropic env gene and the phleo selectable marker conferring resistance to phleomycin (Gatignol et al., FEBS Letters (1988) 230:171-175). By using a PCR-mediated mutagenesis strategy which modifies the end of env gene (see fig. 2), a 76 bp linker was inserted between the stop codon of env and the start  
35 codon of phleo. This allows expression of phleo from the



env mRNA by re-initiation of translation. In addition compared to known env-expressing constructs, this strategy of construction has reduced the length of sequences overlapping with the ends of conventional retroviral vectors. The env genes of Mo-MLV, FeLV, NZB.1V6, 10A1, GALV and RD114 are expressed by plasmids FBdelPMoSAF, FBdelPBSAF, FBdelXSAF, FBdelPGSAF, FBdelp10A1SALF and FBdelPRDSAF, respectively, by using the same backbone as FBdelPASAF (fig. 3). Retroviral vectors produced with the RD114 envelope will be useful for in vivo gene delivery as comparatively to MLV ecotropic or amphotropic envelopes, virions pseudotyped with RD114 envelopes are not inactivated by human complement when they are produced by Mink Mv-1-Lu cells or by some human cells (Table 1).

The HT1080 cell line, isolated from a human fibrosarcoma (ATCC CCL121). The TE671 cell line isolated from a human rhabdomyosarcoma (ATCC CRL 8805) (purchased from ATCC, and tested for absence of usual cell culture contaminants by ECACC), has been used for the definitive construction of packaging cell lines. HT1080 line was chosen among a panel of primate and human lines because MLV-A and RD114 efficiently rescued retroviral vectors from these cells and also because RD114 pseudotypes produced by this cell line were stable when incubated in human serum. In a standard assay (Takeuchi et al., J Virol (1994), 68, 8001-8007), these latter viruses were found more than 500 fold more stable than similar pseudotypes produced in 3T3 cells.

Another advantage for the use of non murine cells to derive packaging lines is the absence of MLV-related endogenous retroviral-like sequences (like VL30 in 3T3 cells) that can cross-package with MLV-derived retroviral vectors (Torrent et al., 1994) and generate potentially harmful recombinant retroviruses.

The helper constructs were introduced into other cell lines (HT1080 (table 2) Mink Mv-1-Lu (table 2), 3T3 (not shown), TE671 (table 2)) for the purpose of comparisons of the efficiency of the constructs.

5 As illustrated hereinafter (Table 2), the reverse transcriptase (RT) activity (provided by expression of the pol gene) in cells transfected with CeB is significantly higher than that of the same cells transfected by the parental plasmid pCRIP or that of cells chronically infected by MLV. This enhancement of viral gene expression is correlated with the titers of lacZ retroviral vectors when an envelope is provided in CeB-lacZ cells after comparison with titers of lacZ pseudotypes of either replication-competent viruses or other helper-free packaging systems.

10 For the generation of final packaging cell lines, the best clonal env transfectants have been selected. Packaging systems obtained in this way will be able to produce helper-free retroviral vectors at titers greater than  $10^8$  infectious particles per ml, which would be 10-100 fold higher to helper-free preparations of others.

25 Because of the way the selectable markers are expressed (see above), growing the packaging cells in phleomycin and blasticidin selective pressure increase and stabilize the expression of the retroviral components and particularly the envelopes, as it is possible that env glycoproteins have toxic effects for the producer cells in the long term which may lead to a decrease of expression.

30 Such an enhancement of viral production observed with the packaging systems described herein might increase the emergence of unwanted retroviruses having recombined between the genomes of both the retroviral vector and either of the

two packaging-deficient constructs. However, the constructs have been designed in such a way that it reduces the probability of emergence of recombinant viruses compared to the parental constructs. To check their safety, attempts have been made to detect the presence of replication-competent retroviruses by a mobilisation assay of a lacZ provirus. No RC viruses have been found in all retroviral vector preparations tested so far.

The following Examples illustrate the invention.

#### Example 1

##### Preparation of Cell lines and viruses.

The following cell lines were used:

A204 (ATCC HTB 82), HeLa (ATCC CCL2), HT1080 (ATCC CCL121), MRC5 (ATCC CCL171), T24 (ATCC HTB 4), VERO (ATCC CCL81) and D17 (ATCC CCL183) were purchased from ATCC.

HOS, TE671 and Mv-1-Lu cells and their clones harboring MFGnslacZ retroviral vector as described by Takeuchi et al., J Virol (1994), 68, 8001-8007.

The above cell lines were grown in DMEM (Gibco-BRL, U.K.) supplemented with 10% fetal calf serum.

EB8 (Battini et al., J. Virol (1992) 66: 1468-1475); psiCRE, psiCRELLZ and psiCRIP (Danos et al., Proc. Natl. Acad. Sci USA (1988) 85: 6460-6464);

Cells GP+EAM12 (Markowitz et al., Virology (1988), 167, 400-406); and

NIH-3T3 murine fibroblasts.

These cell lines were grown in DMEM (GIBCO-BRL, U.K.) supplemented with 10% new-born calf serum.

Mv-1-Lu, TE671 and HT1080 cells were transfected using calcium-phosphate precipitation method (Sambrook et al., "Molecular Cloning" 1989, Cold Spring Harbour Laboratory Press: New York) as described elsewhere (Battini et al., supra). CeB-transfected Mv-1-Lu, TE671 and HT1080 cells were selected with 3, 6-8 and 4  $\mu\text{g/ml}$  of blasticidin S (ICN, UK), respectively, and blasticidin-resistant colonies were isolated 2-3 weeks later. Cells transfected with the various env-expression plasmids were selected with phleomycin (CAYLA, France): 50  $\mu\text{g/ml}$  (for FBASALF-transfected cells) or 10  $\mu\text{g/ml}$  (for FBASAF-, FbdelPASAF-, FbdelPMOSAF, FbdelPIOAISAF or FbdelPRDSAF-transfected cells). Phleomycin-resistant colonies were isolated 2-3 weeks later.

Production of lacZ pseudotypes using replication competent viruses, amphotropic murine leukemia virus (MLV-A) 1504 strain and cat endogenous virus RD114, was carried out as described previously (Takeuchi et al., J Virol (1994), 68, 8001-8007).

## Example 2

### Preparation of Plasmids.

The env gene of pCRIP (Danos et al., supra) was excised by HpaI/ClaI digestion. A 500 bp PCR-generated DNA fragment was obtained using pSV2-bsr (Izumi et al., Experimental Cell Research (1991), 197, 299-233) as template and a pair of oligonucleotides:

(5'>CGGAATTCGGATCCGAGCTCGGCCCCAGCCGGCCACCATGAAAACATTTAACATTTTC TC) (SEQ ID NO 2) at 5' end and

(5'>GATCCATCGATAAGCTTGGTGGTAAACTTTT) (SEQ ID No 3) at 3' end, with SfiI and ClaI sites, respectively. This fragment was inserted in HpaI/ClaI sites of pCRIP by co-ligation with a 85 bp HpaI/SfiI DNA fragment isolated from pOXEnv (Russell et al., Nucleic Acids Research (1993), 21, 1081-1085) which

provides the end of the Moloney murine leukemia virus (MoMLV) pol gene. The resulting plasmid named CeB (Fig. 1) could express the MoMLV gag-pol gene as well as the bsr selectable marker conferring resistance to blasticidin S, both driven by the MoMLV 5'LTR promoter.

A series of env-expression plasmids was generated using the 4070A MLV (amphotropic) env gene (Ott et al., J Virol (1990), 64, 757-766) and the FB29 Friend MLV promoter (Perryman et al., Nucleic Acid Res (1991), 19, 6950). In FBASALF (Fig. 1) a BglIII/ClaI fragment containing the env gene was cloned in BamHI/ClaI sites of plasmid FB3LPh which also contained the C57 Friend MLV LTR driving the expression of the phleo selection marker. A 136 bp env fragment was generated by PCR using plasmid FB3 (Heard et al., J Virol (1991), 65, 4026-4032) as template and a pair of oligonucleotides: (5'>GCTCTTCGGACCCTGCATTC) (SEQ ID NO 4) at 5' end (before ClaI site) and (5'>TAGCATGGCGCCCTATGGCTCGTACTCTATAGGC) (SEQ ID NO 5) at 3' end, providing a KasI restriction site immediately after the env stop codon. This PCR fragment was digested using ClaI and KasI. A DNA fragment containing the FB29 LTR and the MLV-A env gene was obtained by NdeI/ClaI digestion of FBASALF. The fragments were co-ligated in NdeI/KasI digested pUT626 (kindly provided by Daniel Drocourt, CAYLA labs, France). In the resulting plasmid, named FBASAF (Fig. 1), the phleo selectable marker was expressed from the same mRNA as the env gene. A BglIII restriction site was created after the MscI site at position 214 in the FB29 leader by using a commercial linker (Biolabs, France). A NdeI/BglIII fragment containing the FB29 LTR was co-inserted with the BglIII/ClaI env fragment in NdeI/ClaI-digested FBASAF plasmid DNA, resulting in plasmid FBdelPASAF (Fig. 1). Compared to FBASAF, FBdelPASAF has a 100bp larger deletion in the leader region.

## Example 3

## Cloning and Sequencing of the RD114 env gene

The RD114 env gene was first sub-cloned in plasmid Bluescript KS+ (Stratagene) as a 3 Kb HindIII insert isolated from SC3C, an RD114 infectious DNA clone (Reeves et al., J. Virol (1984), 52, 164-171). A 2.7 kb Scal-Hind III fragment of this subclone containing the RD114 env gene was sequenced (Figure 4 (SEQ ID NO 1) - EMBL accession number; X87829). The 5' non-coding sequence upstream of an NdeI site was deleted by an EcoRI/NdeI digestion followed by filling-in with Klenow enzyme and self-ligation. From this plasmid, two DNA fragments were obtained: a BamHI/NcoI 2.5 Kb fragment and a 63 bp PCR-generated DNA fragment using (5'>CGCCTCATGGCCTTCATTAA) (SEQ OD NO 6) at 5' end (before NotI site) and (5'>TAGCATGGCGCCTCAATCCTGAGCTTCTTCC) (SEQ ID NO 7) at 3' end, providing a KsI restriction site just after RD114 env gene stop codon. The PCR fragment was digested with NcoI and KsI. Both fragments were co-inserted between BglII and KsI sites of FBdelPASAF and the resulting plasmid was named FBdelPRDSAF (Fig. 1).

Plasmid pCRIPAMgag- (Danos, O. et al., Proc Natl Acad Sci USA (1988) 85:6460-6464) was used for transfection.

## Example 4

## Infection assays.

Target cells were seeded in 24-multiwell plates ( $4 \times 10^4$  cells per well) and were incubated overnight. Infections were then carried out at 37°C by plating 1 ml dilutions of viral supernatants in the presence of 4 µg/ml polybrene (Sigma) on target cells. 3h later virus-containing medium was replaced by fresh medium and infected cells were incubated for two days before X-gal staining, performed as previously described (Tailor et al., J Virol (1993), 67, 6737-6741,

Takeuchi et al., J Virol (1994), 68, 8001-8007). Viral titers were determined by counting lacZ-positive colonies as previously described (Cosset et al., J. Virol. (1990) 64: 1070-1078). Stability of lacZ pseudotypes in fresh human serum was examined by titrating surviving virus after incubation in 1:1 mixture of virus harvest in serum-free medium and fresh human serum for 1 h at 37°C as described before (Takeuchi et al. supra).

#### Example 5

##### Reverse transcriptase (RT) assay.

RT assays were performed either as described previously (Takeuchi et al. supra) or using an RT assay kit (Boehringer Mannheim, U.K.) following the manufacturer's instruction but using MnCl<sub>2</sub> (2 mM) instead of MgCl<sub>2</sub>.

#### Example 6

##### Screening producer cell lines.

Viral particles generated with RD114 envelopes have been found to be more stable in human serum than virions with MLV-A envelopes and that the producer cell line also controls sensitivity (Takeuchi et al. supra). A panel of cell lines was screened for their ability to produce high titer viruses and for the sensitivity of these virions to human serum. To do this, cells were infected at high multiplicity with lacZ pseudotypes of either MLV-A or RD114 and cells producing helper-positive lacZ pseudotypes were established. Human HT1080 and TE671 and mink Mv-1-Lu cells were found to release high titer lacZ(RD114) and lacZ(MLV-A) viruses. LacZ(MLV-A) pseudotypes produced by HT1080 cells were more resistant to human serum than those produced by other cells. The titer of these viruses was only four-fold less following a 1 hr incubation with human serum than a

control incubation (Table 1). LacZ(RD114) pseudotypes produced by human cells or mink Mv-1-Lu cells were in general stable in human serum (Table 1). These results suggested that HT1080, TE671 and Mv-1-Lu cells provided the best combination of high lacZ titers and resistance to human serum and they were therefore used for the generation of retroviral packaging cells.

Table 1. Titer and stability of lacZ pseudotypes.

Producer cell	LacZ (MLV-A)		LacZ (RD114)	
	Titer <sup>a</sup>	Stability <sup>b</sup>	Titer <sup>a</sup>	Stability <sup>b</sup>
A204	650	<3	1,200	105
HeLa	9	nd	2,000	115
HOS	4,500	6	23,000	86
HT1080	2,000,000	26	400,000	129
MRC-5	450	10	1,000	nd
T24	350	nd	1,200	nd
TE671	15,000	2	90,000	38
VERO	260	nd	90	nd
D17	900	<1	200,000	1
Mv-1-Lu	80,000	1	200,000	120

a: titration on TE671 cells as lacZ i.u./ml

b: % of infectivity of human serum-treated viruses compared to fetal calf serum-treated viruses

#### Example 7

Construction of an improved gag-pol expression vector.

A MoMLV gag-pol expression plasmid, CeB (Fig. 1), was



derived from pCRIP (Danos et al., Proc. Natl. Acad Sci USA (1988) 85: 6460-6464). Approximately 2 Kb of env sequence were removed from pCRIP and the bsr selectable marker, conferring resistance to blasticidin S (Izumi et al., Experimental Cell Research (1991) 197:229-233), was inserted 74 nts downstream of the gag-pol gene. This 74 nts interval had no ATG triplets and was thought to provide an optimal distance between the stop codon of the pol reading frame and the start codon of the bsr gene to allow re-initiation of translation (Kozak Mol Cell Biol., 1987, 7: 3438-3445). There was no "Kozak" consensus sequence (Kozak Cell, (1986) 44: 283-292) at the 5' end of the marker gene. Therefore, bsr could only be expressed by re-initiation of translation after the upstream gag-pol gene had been expressed. Consequently, after transfection of CeB in Mv-1-Lu/MFGnlsLacZ (ML), TE671/MFGnlsLacZ (TEL) or HT1080 cells, blasticidin S-resistant bulk populations and most cell clones expressed high levels of gag-pol proteins assessed by the reverse-transcriptase (RT) activity found in cell supernatants (Table 2). Considerably higher RT activities were found in bulk populations of CeB-transfected ML cells compared to bulk population of ML cells stably transfected with the parental pCRIP construct. Similarly the RT activities of two packaging cell lines generated using pCRIPenv- construct, psiCRE cells (Danos et al., supra) and EB8 cells (Battini supra.) were less than that of CeB transfected clones (Table 2). Finally, RT activity in CeB transfected cell supernatants was higher than that of cells chronically infected by replication-competent MLV-A (Table 2).

Table 2. Secreted reverse transcriptase expression

Cell <sup>a</sup>	RT activity <sup>b</sup>	LacZ Titer <sup>c</sup>
-------------------	--------------------------	-------------------------

24

	ML/MLV-A	1	$8 \times 10^4$
	MLSvB	0.1	<1
	MLCRIP (bulk)	0.15	nd
	MLCeB (bulk)	1.7	nd
5	MLCeB1	4.2	$1 \times 10^6$
	MLCeB4	1.6	$1 \times 10^6$
	TEL/MLV-A	3.6	$2 \times 10^6$
	TELCeB6	5.2	$4 \times 10^7$
	HT1080/MLV-A	1.1	$1 \times 10^6$
10	HTCeB6	1.9	$1 \times 10^6$
	HTCeB18	2.7	$2 \times 10^6$
	HTCeB22 (FLY)	6.9	$5 \times 10^6$
	HTCeB48	5.5	$3 \times 10^6$
	EB8	0.22	$1 \times 10^4$
15	psiCRE-LLZ	1.2	$1 \times 10^{5d}$

a: ML, Mv-1-Lu cells harboring a MFGnslacZ provirus; TEL, TE671 cells harboring a MFGnslacZ provirus; /MLV-A, cells chronically infected with MLV-A 1504 strain; MLSvB, ML cells transfected with a plasmid pSV2bsr alone; MLCRIP, ML

cells co-transfected with pCRIP and pSV2bsr.

b: Average of arbitrary units relative to ML/MLV-A RT activity of at least two independent experiments was shown. The standard errors did not exceed 20 % of the values.

c: titration on TE671 cells as lacZ i.u./ml. After polyclonal transfection of a plasmid which expresses MLV-A env in MLCeB clones, TELCeB clones, HTCeB clones and EB8 cells; nd, not done.

d: titration on NIH3T3 cells

To rescue infectious lacZ viruses, MLCeB and TELCeB clones were transfected with FBASALF DNA, a plasmid designed to express the MLV-A env gene (Fig. 1). Bulk populations of stable FBASALF transfectants were isolated and supernatants were titrated using TE671 cells as targets. Titers of lacZ viruses were higher than either MLV-A infected ML or TEL cells, or FBASALF-transfected EB8 cells (Table 2). These data suggested that CeB was an extremely efficient MLV gag-pol expression vector in mink Mv-1-Lu and TE671 cells. CeB

was therefore used to derive packaging cells by transfection of HT1080 cells. 41/49 blasticidin S-resistant colonies had detectable levels of RT; 9 had RT activity higher than that of control MLV-A-infected HT1080 cells (data not shown).

5 Expression of gag precursor was confirmed in cell lysates and supernatants of these 9 HTCeB clones by immunoblotting using antibodies against p30-CA (data not shown). The 4 clones with the highest expression of gag proteins (clones 6,18,22 and 48) were infected at high-multiplicity with  
10 helper free, lacZ pseudotypes bearing MLV-A envelopes (MFGn1slacZ(A)) produced by TELCeB6/FBASALF (Table 3) and then transfected with FBASALF. Supernatants of bulk, phleomycin-resistant transfectants were assessed for RT activity and lacZ titer (Table 2). Clone HTCeB22, named FLY,  
15 was found to be the best gag-pol producer clone and was used to introduce env expression vectors for the generation of packaging cell lines.

Table 3. Titer following env construct transfection

	Producer cell	Env source	Titer <sup>a</sup>
5	psiCRIP lacZ 5	pCRIPAMgag-	6x10 <sup>4b</sup>
	GP+EAM12 lacZ 25	envAM	3x10 <sup>5b</sup>
10	TELCeB6	FBASALF <sup>c</sup>	5x10 <sup>7</sup>
		FBASAF <sup>c</sup>	2x10 <sup>7</sup>
		FbdelPASAF <sup>c</sup>	2x10 <sup>7</sup>
15	TELCeB6	FbdelPASAF 1	3x10 <sup>7</sup>
		FbdelPASAF 4	2x10 <sup>7</sup>
		FbdelPASAF 6	1x10 <sup>7</sup>
		FbdelPASAF 7	5x10 <sup>7</sup>
		FbdelPASAF 8	1x10 <sup>7</sup>
20		FbdelPRDSAF 2	1x10 <sup>6</sup>
		FbdelPRDSAF 4	3x10 <sup>5</sup>
		FbdelPRDSAF 7	1x10 <sup>7</sup>
		FbdelPRDSAF 8	2x10 <sup>6</sup>
25	FLY <sup>d</sup>	FbdelPASAF 1	1x10 <sup>1</sup>
		FbdelPASAF 4	1.5x10 <sup>6</sup>
		FbdelPASAF 5	1x10 <sup>6</sup>
		FbdelPASAF 7	1x10 <sup>6</sup>
		FbdelPASAF 13	7x10 <sup>6</sup>
		FbdelPASAF 14	4x10 <sup>6</sup>
30		FbdelPASAF 15	1x10 <sup>6</sup>
		FbdelPASAF 16	5x10 <sup>6</sup>
		FbdelPASAF 17	6x10 <sup>6</sup>
35	FLYA4 lacZ 3	FbdelPASAF 4	2x10 <sup>7b</sup>
	FLY <sup>d</sup>	FbdelPRDSAF 1	2.5x10 <sup>6</sup>
		FbdelPRDSAF 2	1x10 <sup>7</sup>
		FbdelPRDSAF 6	5x10 <sup>6</sup>
40		FbdelPRDSAF 10	2x10 <sup>6</sup>
		FbdelPRDSAF 11	3x10 <sup>6</sup>
		FbdelPRDSAF 13	1x10 <sup>6</sup>
		FbdelPRDSAF 17	5x10 <sup>6</sup>
		FbdelPRDSAF 18	3x10 <sup>7</sup>
45		FbdelPRDSAF 19	6x10 <sup>6</sup>

Average titers of at least three independent experiments were shown. The standard errors did not exceed 30 % of the titer values.

a: titrated on TE671 cells as lacZ i.u./ml

b: results of best MFGnslacZ producer clones.

c: bulk populations of env-transfectants in TELCeB6 cells.

d: titration after bulk infection with helper-free MFGnslacZ.

5      **Example 8**

Construction of env expression vectors.

A series of MLV-A env expression plasmids were then generated (Fig. 1). In FBASALF, the env gene was inserted  
10      between two Friend-MLV LTRs, its expression driven by the FB29 MLV LTR (Perryman et al., supra). Most of the packaging signal located in the leader region was deleted. This plasmid also expressed the phleo selectable marker (Gatignol et al., supra) driven by the 3' LTR. FBASAF and FBdelPASAF  
15      were then designed following the same strategy used for CeB. These two vectors differed only by the extent of deletion of the packaging signal, FBdelPASAF having virtually no leader sequence. Compared to pCRIPAMgag- and pCRIPgag-2 env plasmids expressed in psiCRIP or psiCRE packaging cells  
20      (Danos et al., supra) about 5 Kb of gag-pol sequences was removed. In addition the 258 bp retroviral sequence containing the end of env gene and the beginning of U3 found in pCRIPAMgag- and pCRIPgag-2 was also removed. For both FBASAF and FBdelPASAF plasmids, the phleo selectable marker  
25      was inserted downstream of the env gene by positioning a 76 nts linker with no ATG codons between the two open-reading frames. Phleo could therefore only be expressed by re-initiation of translation by the same ribosomal unit that had expressed the upstream env open reading frame.  
30      FBdelPASAF was also used to generate FBdelPRDSAF, an RD114 envelope expression plasmid (Fig. 1).

After transfection of the env plasmids into TELCeB6 cells (Table 2), bulk populations of phleomycin-resistant colonies  
35      were isolated and their production of lacZ virus measured

(Table 3). FBASALF gave a titer of  $5 \times 10^7$  lacZ-i.u./ml, whilst titers with either FBASAF or FBdelPASAF were  $2 \times 10^7$  lacZ-i.u./ml (Table 3). Titers of  $5 \times 10^7$  or  $10^7$  lacZ-i.u./ml could be obtained with some FBdelPASAF cell clones or FBdelPRDSAF clones, respectively.

As FBdelPASAF has minimal virus-derived sequences and was shown to be the safest construct (see below and Table 4), it and FBdelPRDSAF were used to generate packaging lines from FLY cells (clone HTCeB22, Table 2). Envelope expression of these clones was assayed by interference to challenge with MFGnslacZ(A) or MFGnslacZ(RD) pseudotypes produced by TELCeB6/FBdelPASAF-7 or TELCeB6/FBdelPRDSAF-7, respectively (Table 3). The cell lines showing most interference were cross-infected at high multiplicity with these pseudotypes to provide MFGnslacZ proviruses, and supernatants were then titrated on TE671 cells (Table 3). FLY-FBdelPASAF-13 (FLYA13 packaging line) and FLY-FBdelPRDSAF-18 (FLYRD18 packaging line) gave the highest productions of lacZ viruses, around  $10^7$  lacZ-i.u./ml. The best MFGnslacZ producer clones derived from either psiCRIP cells (Danos et al., supra) or GP+EAM12 cells (Markowitz et al., supra) gave approximately 50 fold lower titers (Table 3). The lacZ titers of the FLY-derived lines shown in Table 3 are lower than the best TELCeB6-derived lines after transfection of either FBdelPASAF or FBdelPRDSAF (Table 3). However it should be noted that the lacZ provirus expressed in TELCeB6 cells was obtained after clonal selection but was introduced polyclonally in FLY-derived env-transfected cell clones. When FLY-FBdelPASAF-4 cells (FLYA4 packaging line), infected with helper-free MFGnslacZ(RD), were cloned by limiting dilution the best clones (eg. FLYA4lacZ3) were found to produce 20 times more infectious viruses than the bulk population, reaching the range of titers obtained with the best TELCeB6-FBdelPASAF clones (Table 3).

## Example 9

Assays for transfer of gag-pol or env functions.

To assay for replication-competent viruses, supernatants  
5 were used to infect TEL cells (a clone of TE671 cells  
harboring an MFGnls lacZ provirus). Infected cells were  
passaged for 6 days or longer and their supernatants were  
used for infection of fresh TE671 cells. No transmission of  
lacZ viruses could be detected (Table 4), demonstrating that  
10 the supernatants of pCRIPAMgag--, FBASALF-, FBASAF-, or  
FBdelpASAF-transfected TELCeB6 cells were helper-free.  
Similar absence of replication competent recombinant  
retroviruses was demonstrated using supernatant from a  
clone of psiCRIP-MFGnls lacZ cells or from two clones of  
15 FLYA-MFGnls lacZ cells (Table 4).

There have been reports that helper-free retroviral vector  
stocks may nevertheless contain recombinant retroviruses  
(replication incompetent) carrying either gag-pol or env  
20 genes (Bestwick et al., Proc Natl Acad Sci USA (1988), 85,  
5404-5408, Cosset et al., Virology (1993), 193, 385-395,  
Girod et al., Virology (1995), in press). To assay for such  
recombinant retroviruses, mobilisation of an MFGnls lacZ  
provirus from two indicator cell lines which could cross-  
25 complement potential recombinant viruses carrying either  
gag-pol or env functional genes was attempted. The TELCeB6  
line (Table 2) expressing gag-pol proteins was used as  
indicator cell line to test for the presence of env  
recombinant (ER) viruses. The TELMOSAF indicator line  
30 expressing MoMLV env glycoproteins (obtained by transfection  
of FBMOSAF, a plasmid expressing the MoMLV env gene using  
FBASAF backbone, in TEL cells) was used to detect the  
presence of gag-pol recombinant retroviruses (GPR viruses).  
After passaging 4-8 days, the supernatants of the infected  
35 indicator cells were used to infect either human TE671 cells

or murine NIH3T3 cells.

5 TELCeB6 cells transfected with various env-expressing constructs, pCRIPAMgag-, FBASAF and FBdelPASAF were compared. Although the supernatants of TELCeB6-FBdelPASAF cells were devoid of replication-competent retroviruses, they were found sporadically to transfer gag-pol genomes (Table 4). No GPR viruses could be detected when less than  $2 \times 10^5$  virions were used to infect the indicator cells.

10 Similarly TELCeB6 indicator cells infected with various helper-free viruses were shown sporadically to release lacZ virions (Table 4). The number depended both on the env-expression vector used and on the virus input quantity. Compared to lacZ viruses generated using pCRIPAMgag-

15 plasmid, the frequency of detection of the env-recombinant viruses was lower for supernatants generated by using FBASAF and FBdelPASAF constructs (Table 4). For FBdelPASAF construct when less than  $5 \times 10^5$  MFGnslacZ(A) helper-free virions were used to infect the indicator cells, no ER

20 retroviruses could be detected. From these experiments, it could be estimated that a supernatant, produced from TELCeB6-FBdelPASAF cells, containing  $1 \times 10^7$  infectious units of MFGnslacZ retroviral vector contained no replication-competent virus, and about 100 gag-pol and 100 env

25 recombinant retroviruses.



Table 4. Transfer of packaging function

	Producer cell	Indicator cell	Input virus <sup>a</sup> (lacZ-i.u.)	Detection <sup>b</sup>		
				++	+	-
5	Replication competent virus					
	psiCRIP lacZ 5	TEL	2x10 <sup>4</sup>	0/4	0/4	4/4
10	TELCeB6-pCRIPAMgag-	TEL	5x10 <sup>6</sup>	0/4	0/4	4/4
	TELCeB6-FBASAF	TEL	5x10 <sup>6</sup>	0/4	0/4	4/4
	TELCeB6-FBdelPASAF	TEL	5x10 <sup>6</sup>	0/4	0/4	4/4
15	FLYA4 lacZ 3	TEL	1x10 <sup>7</sup>	0/4	0/4	4/4
	FLYA4 lacZ 7	TEL	1x10 <sup>7</sup>	0/4	0/4	4/4
Gag-pol recombinant						
20	TELCeB6-FBdelPASAF 7	TELMOSAF	2x10 <sup>7</sup>	0/4	1/4	3/4
	TELCeB6-FBdelPASAF 7	TELMOSAF	2x10 <sup>6</sup>	0/4	2/4	2/4
	TELCeB6-FBdelPASAF 7	TELMOSAF	2x10 <sup>5</sup>	0/4	2/4	2/4
	TELCeB6-FBdelPASAF 7	TELMOSAF	2x10 <sup>4</sup>	0/4	0/4	4/4
Env recombinant						
25	TELCeB6-pCRIPAMgag-	TELCeB6	5x10 <sup>6</sup>	2/4	1/4	1/4
	TELCeB6-pCRIPAMgag-	TELCeB6	5x10 <sup>5</sup>	1/4	1/4	2/4
	TELCeB6-pCRIPAMgag-	TELCeB6	5x10 <sup>4</sup>	0/4	2/4	2/4
30	TELCeB6-FBASAF	TELCeB6	5x10 <sup>6</sup>	0/4	2/4	2/4
	TELCeB6-FBASAF	TELCeB6	5x10 <sup>5</sup>	0/4	1/4	3/4
	TELCeB6-FBASAF	TELCeB6	5x10 <sup>4</sup>	0/4	1/4	3/4
35	TELCeB6-FBdelPASAF	TELCeB6	5x10 <sup>6</sup>	0/4	1/4	3/4
	TELCeB6-FBdelPASAF	TELCeB6	5x10 <sup>5</sup>	1/4	3/4	0/4
	TELCeB6-FBdelPASAF	TELCeB6	5x10 <sup>4</sup>	0/4	0/4	4/4

a: number of lacZ i.u. used to infect indicator cells

b: number of incidence out of four experiments. The ranges of lacZ titers rescued from infected indicator cells are shown for each virus input: >100 lacZ i.u./ml (++) , 1-100 lacZ i.u./ml (+) and <1 lacZ i.u./ml (-).

Titers were determined on TE671 cells for replication competent virus and env recombinant and NIH3T3 cells for

gag-pol recombinant.

#### Example 10

5 In order to confirm resistance to complement and absence of replication competent virus in our best packaging lines, MFGnslacZ(A) and (RD) harvested from FLYA13 and FLYRD18, respectively, after polyclonal transduction of MFGnslacZ (Table 3 above) were tested for stability in fresh human serum and generation of replication competent virus. Titters  
10 of MFGnslacZ(RD) from FLYRD18 after 1 hr incubation with 3 independent samples of fresh human serum were 80 to 120 % of control incubations, while titers of MFGnslacZ(A) from FLYA13 were 50 to 90 % of controls (data not shown). No replication competent virus was detected in the same assay  
15 described above (Table 4) when  $1 \times 10^7$  i.u. each of MFGnslacZ(A) and (RD) were tested.

#### EXAMPLE 11.

##### 20 Generation of plasmids.

CeB plasmid (Fig. 5) expressing MoMLV gag-pol gene, was further modified to remove the splice donor site located in the leader region. A 272 bp fragment was PCR-generated by using OUSD- (5'-TCTCGCTTCTGTTCGCGCGC) and OLSD-  
25 (5'-TCGATCAAGCTTGCGCCGCGGTGGTGGGTGGTCC) as primers and further digested with BssHII and HindIII. A 1008 bp HindIII-XhoI fragment isolated from CeB (encompassing a part of leader sequence and beginning MoMLV gag) and the PCR fragment were co-inserted into pCeB from which the 1275 bp  
30 BssHII-XhoI fragment (encompassing R-U5-leader-gag) had been removed. The resulting plasmid, named pCeB DS- (Fig. 5), beared the deletion of splice donor (SD) site and a NotI restriction site created just downstream to the lost SD site.

A series of gag-pol expression plasmids in which the MoMLV LTR promoter was replaced by the human cytomegalovirus immediate early promoter (hCMV promoter) was derived from both CeB DS- and hCMV-G (Yee et al., 1994 PNAS, 91: 9564-9568), a plasmid used as a source for the hCMV promoter. A NotI-filled/EcoRI 7260 bp fragment was isolated from CeB DS- and cloned into hCMV-G which had been opened with SalI (further rendered blunt-ended) and EcoRI to remove the VSV-G gene. The resulting plasmid was cutted with ClaI and EcoRI to remove a 1155 bp fragment encompassing sequence derived from 3'-LTR and SV40 polyA sequence and self-ligated after filling both protruding DNA ends. The resulting plasmid, named phCMV-intron (Fig. 5), had gag-pol and bsr ORFs inserted between the CMV promoter and rabbit beta-globin polyA post-transcriptional regulatory sequences.

An intermediate plasmid was generated by sub-cloning a 7260 bp EcoRI fragment (isolated from CeB DS-) into hCMVG opened with EcoRI. A 1155 bp fragment (encompassing sequence derived from 3'-LTR and SV40 polyA sequence) was removed from this intermediate plasmid which was then re-circularized by self ligation after filling both ends. The resulting plasmid, named phCMV+intron 2P (Fig. 5), was digested with NotI and the vector was treated with klenow enzyme. A 1440 bp fragment (encompassing hCMV promoter and rabbit beta-1 globin intron B (Rohrbaugh et al., 1985 Mol. Cell Biol, 5: 147-160)) was isolated from phCMV+intron 2P by NotI/EcoRI digestion. This fragment was further treated with klenow enzyme and ligated back into the vector. The resulting plasmid, named hCMV+intron (Fig. 5), could express gag-pol and bsr genes driven by the hCMV promoter and beared an intron sequence derived from rabbit beta-1 globin intron B having both SD and SA (splice acceptable) sites.

A 2450 bp fragment was removed from phCMV+intron 2P by

NotI/XhoI digestion. The resulting vector fragment was then used to co-ligate a 1330 bp fragment (containing hCMV promoter + 5' end of rabbit beta-1 globin intron B (with SD site)) isolated from phCMVG by ApaI-filled/NotI digestion and a 1 kb fragment isolated from phCMV+intron 2P by NotI-filled/XhoI digestion. Compared to phCMV+intron 2P, the resulting plasmid, named hCMV+SD intron (Fig. 5), had the deletion of the 3' end of the rabbit beta-1 globin intron B and thus no SA site in the leader region.

Construct phCMV+leader (Fig. 5) has been described elsewhere (Savard et al., unpublished). This plasmid, in which gag-pol and bsr genes were driven by the hCMV promoter, had the MoMLV SD site in the leader region.

#### Gag-pol expression.

The different constructs, including the parental CeB plasmid, were analysed comparatively in a complementation assay after transfection in TEL-FBdelPASAF cells expressing 4070A-MLV (amphotropic) envelope and harboring a MFGnslacZ provirus. The transient production of lacZ retroviruses as well as the stable production of lacZ retroviral vectors after selection with blasticidin S were determined (Table 5). All the constructs were able to rescue infectious lacZ retroviruses indicating the expression of gag-pol proteins after transient transfection. Most likely due to the efficient hCMV and rabbit beta-1 globin intron B (post)-transcriptional regulatory sequences, hCMV+intron was particularly potent in transient retroviral vector production. However, 10 times less blasticidin-resistant colonies were obtained with hCMV+intron comparatively to CeB, and stable lacZ virus production from hCMV+intron was about 5-10 times lower than that of CeB. Clonal examination of lacZ retrovirus production from blasticidin-resistant colonies indicated that 80-90% of colonies could express

high levels of gag-pol proteins for both hCMV+intron and CeB plasmids. In contrast, despite variation in their ability to form blasticidin-resistant colonies after transfection and despite their ability to express gag-pol proteins from transient transfectants, all other constructs had a weak capacity for rescuing lacZ retroviral vectors from stable transfectants (Table 5).

Table 5. Comparative study of gag-pol-bsr plasmids.

gag-pol-bsr plasmid	Transient (lacZ i.u./ml)	no clones bsr <sup>+</sup>	Stable (lacZ i.u./ml)	% gag-pol /bsr
Ceb	300/ml	50	10 <sup>7</sup>	90%
Ceb DS-	144/ml	5	10 <sup>5</sup>	50%
hCMV+intron 2P	ND	20	10 <sup>6</sup>	50%
hCMV-intron	812/ml	0	-	-
hCMV+SD intron	150/ml	1000	10 <sup>2</sup>	nd
hCMV+leader	328/ml	1000	10 <sup>2</sup> -10 <sup>3</sup>	nd
hCMV+intron	12000/ml	5	10 <sup>6</sup> -10 <sup>7</sup>	80%

Northern blot analyses were performed on stable transfectants (blasticidin-resistant) obtained with some of the gag-pol-bsr plasmids. As expected, the results (not shown) displayed a correlation between expression of gag-pol mRNAs and gag-pol protein expression detected by rescue analysis (Table 5). CeB construct was found to produce 2-3 fold more gag-pol mRNAs compared to hCMV+intron. Interestingly, an unexpected 2.45 kb RNA band was found for hCMV+intron construct at a ratio of 2:1 compared to the abundance of the gag-pol mRNA band (at 5.95 kb). Further

investigations by using other probes revealed that a cryptic splice donor (SD) site located in the gag gene (right in the middle of the CA coding region at position 1596-1597 -numbering according to Shinnick et al., 1981 Nature (London) 293: 543-548) was activated in this latter construct. The 2.45 RNA species, lacking the 3' half of the gag gene and most of the pol gene, is unlikely to give rise to any useful translational product. It is therefore interesting to notice that hCMV+intron construct was able to give rise to slightly more transcripts (gag-pol 5.95 mRNA + 2.45 alternative RNA band) compared to gag-pol mRNA expressed from CeB construct. Therefore we decided to inactivate the cryptic SD site in the hCMV+intron construct in order to increase the ratio of gag-pol mRNAs.

#### Assays for transfer of gag-pol functions.

Although the supernatants of packaging cell lines generated with CeB gag-pol expression construct were devoid of replication-competent retroviruses, they were found sporadically to transfer gag-pol genomes (example 9, Table 4) (Cosset et al., 1995 J. Virol 69: 7430-7436). Because gag-pol-bsr constructs generated here by using the hCMV promoter had much less retroviral sequences homologous to the retroviral vector than the parental CeB construct (Fig. 5), they are less likely to give rise to gag-pol recombinant (GPR) viruses. Therefore, the most efficient gag-pol-bsr plasmids, hCMV+intron and CeB, were further analysed for emergence of GPR viruses. To assay for such recombinant retroviruses, we attempted to mobilise an lacZ provirus from an indicator cell lines which could cross-complement potential recombinant viruses carrying gag-pol functional genes. Results displayed in Table 6 showed that consistently with data reported previously (example 9, Table 4) (Cosset et al., 1995 Supra), lacZ retrovirus vectors generated by using CeB gag-pol construct were contaminated with GPR viruses. In

contrast lacZ retrovirus vectors generated by using hCMV+intron construct were completely devoid of such GPR viruses, suggesting that this construct was improved compared to CeB with respects with emergence of recombinant viruses.

Table 6. Comparative study of gag-pol-bsr plasmids.

plasmid	input virus (lacZ i.u.) <sup>a</sup>	no of experiments giving titres of <sup>b</sup>		
CeB	5x10 <sup>6</sup>	5	3	0
	5x10 <sup>5</sup>	2	4	2
	5x10 <sup>4</sup>	0	1	7
hCMV+intron	5x10 <sup>6</sup>	0	0	8
	5x10 <sup>5</sup>	0	0	8
	5x10 <sup>4</sup>	0	0	8

4x10<sup>4</sup> cells of TEL/MOSAF in 24 wells were challenged with lacZ(A) of i.u. indicated in the table (a), and incubated at 37°C for 3 days. Cells were trypsinized and transferred into small flasks. Cell sup was harvested on day 5 after lacZ(A) challenge and plated on either TE571 (not shown) and 3T3 cells (b). No lacZ was mobilized into TE671 at all. LacZ(A) from CMV-int 10 again did not rescue lacZ from TEL/MOSAF.

#### Example 12

Generic primers to detect D-type (Medstrand and Blomberg J.Virol. (1993) 67:6778-6787), C-type (Shih et al., J Virol. (1989) 63:64-75), human endogenous virus RTVL-H (Wilkinson et al., J.Virol. (1993) 67:2981-2989), by RT-PCR were employed (Patience et al., supra). Primers to detect mouse endogenous VL30 element (Adams et al Mol.Cel.Biol. (1988) 8:2989-2998), and MFGnslacZ RNA were designed and synthesized (TABLE X). Overnight supernatants (in 4ml of culture medium) from 106 cells of GP+EAM12lacZ25, FLYA4lacZ3

and TELCeB6FBASALF cells (Table 3) were harvested and centrifuged in sucrose gradient as described previously (Patience et al., J.Virol., 70:2654-2657). Fractions containing retrovirus particles were collected, and RNA extracted. One twentieth of the RNA preparation or dilution's thereof were applied to RT-PCR as described previously (Table X). A 1/200 of RNA harvested from GP+EAM12lacZ25 cells was positive for VL30 RNA. MFGnls lacZ RNA was found from 1/20 of RNA from GP+EAM12lacZ and TELCeB6FBASALF cells and 1/200 of RNA from FLYA4lacZ3 cells. The primer combinations for RTVL-H, C- and D-type RNA did not give detectable PCR product.

Table 7. RT-PCR detection of endogenous retrovirus RNA associated with virus particles.

RNA	primer (5'-3') forward(F)/reverse(R)	rt-pcr of virion associated RNA from <sup>a</sup>		
		GP+EAM12 lacZ25	FLYA4 lacZ3	TELCeB6F BASALF
MFGnls lacZ	F) CTCTGGCTCACAGTACGACGTAG R) CCATCAATCCGGTAGGTTTTCCG	+	++	+
C-type	F) CARRGKTTCAARAACWSYCCCAC R) AGYARVGTAGCNGGGTTHAGG	-	-	-
D-type	F) TCCCCTTGGAATACTCCTGTTTTYGT R) CATTCCTTGTGGTAAACTTTCCAYTG	-	-	-
RTVL-H	F) CCTCACCTGATCACRYTTG R) GAATTATGTCTGACAGAAGGG	NT	-	-
VL30	F) GTTGACATCTGCAGAGAAAGACC R) TCTGAGGTCTGTACACACAATGG	++	NT	NT



-----  
a:-, not detected; + detected in 1/20 RNA preparation; ++ detected in 1/200 RNA preparation; NT, not tested because the cells do not possess the corresponding genes.

5

**EXAMPLE 13.****Generation of gag-pol pre-packaging cells by using TE671 cells.**

10 CeB, a plasmid designed to over-express MoMLV gag and pol proteins was introduced in TE671 human rhabdomyosarcoma cells (ATCC CRL8805). After selection with blasticidin, 50 bsr-positive colonies were isolated and the RT (reverse transcriptase) activity was analysed in their supernatants.

15 12 TE671-CeB (TECeB) clones with high RT activity were selected for further analysis. The best TECeB clone, clone #15, had a RT activity roughly equivalent to that TELCeB6 cells (Cosset et al., J. Virol. 69:7430-7436 (1995); see also Example 7, Table 6 in this patent application) but

20 displayed 2-3 fold more gag-precursors into cells as demonstrated in immunoblots by using anti-CA antibodies. The biological activity of gag-pol proteins expressed in the six best TECeB clones was further confirmed by their ability to produce infectious retroviruses in a complementation assay.

25 A lacZ provirus was introduced into each of the TECeB clones by polyclonal cross-infection by using lacZ(RD114) helper-free retrovirus vectors. FBMOSALF, a MoMLV env expression plasmid (Cosset et al., J. Virol. 69:6314-6322), was then transfected in each of the TECeB-lacZ lines and in the

30 TELCeB6 cell line for comparison. After selection with phleomycin, the titer of lacZ retrovirus vectors was determined in the supernatant of pools of phleomycin-resistant colonies for each TECeB-lacZ-FBMOSALF lines. A

good correlation was found between gag-pol expression into the TE-CeB clones (as determined by RT-assays and anti-gag immunoblots) and their ability to release infectious lacZ particles. TE-CeB15 cells could release approximately the same number of lacZ particles when compared to TELCeB6 cells although TELCeB6 cells had the advantage of being selected for lacZ expression (Cosset et al., J. Virol. 69:7430-7436 (1995)). TE-CeB15 cells were therefore used to derive retroviral packaging cell lines.

#### Construction of env-expression plasmids.

A series of plasmid (Fig. 3) was designed to allow expression of different retroviral envelope genes (isolated from MoMLV, GALV -Gibbon Ape Leukemia Virus-, and MLV-10A1). FBdelPMOSAF (Fig. 3, nucleotide sequence in Fig. 10) and FBdelP10A1SAF, expressing ecotropic MoMLV or MLV-10A1 envelopes, were generated by replacing the BglIII/ClaI fragment from FBdelPASAF (Cosset et al., J. Virol. 69:7430-7436 (1995); see also Example 7, Fig. 2 and nucleotide sequence in Fig. 9) encompassing most of the env gene and splice acceptor site with that of MoMLV (position 5407 to 7679, Shinnik et al., 1981) or with that of MLV-10A1 (Ott et al., J. Virol. 64:757-766 (1990)). Nucleotides 7514-7516 of GALV (Delassus et al., Virology 173:205-213 (1989)) were mutated by PCR-mediated mutagenesis to create a ClaI site (AAG to CGA), thereby introducing a conservative modification (a lysine (amino-acid 665 of GALV env precursor) to an arginine). The BamHI/ClaI fragment (nts 4994 (Delassus et al. Virology 173:205-213 (1989)) to 7517) was then sub-cloned into FBdelPASAF in which the BglIII/ClaI encompassing most of the env gene and splice acceptor site had been removed. The resulting plasmid, expressing GALV

envelope glycoproteins, was named FBdelPGASAF (Fig. 3, nucleotide sequence in Fig. 11).

CMV10A1 was generated by inserting a Klenow enzyme-filled EagI/SalI fragment from FBdelP10A1SAF (encompassing 10A1 MLV env gene and phleo selectable marker) into hCMV-G digested with BamHI and filled with Klenow enzyme. The resulting plasmid, CMV10A1 (Fig. 3 and nucleotide sequence in Fig. 13) could express 10A1 envelopes under control of the hCMV promoter and the phleo selectable marker by translation re-initiation.

**Generation of a multi-tropic set of TE671-based retroviral packaging lines.**

FBdelPRDSAF (Fig. 3, nucleotide sequence in Fig. 12), FBdelPASAF, FBdelPGASAF, FBdelPMOSAF and FBdelP10A1SAF were independently introduced into cells of the TE-CeB15 pre-packaging line, expressing MoMLV gag-pol proteins.

Transfected cells were phleomycin-selected and 15-20 phleo-resistant colonies were isolated for each env-expression plasmid transfected.

Individual colonies were then analysed for expression of envelope glycoproteins by immunoblots on cell lysates by using antibodies against RD114 SU glycoproteins or against Rausher leukemia virus SU (to screen MoMLV, MLV-4070A and MLV-10A1 env-producer clones) or against GALV. The best env-producer colonies as determined in this assay were further analysed by a complementation assay after introducing a lacZ retroviral vector. LacZ pseudotypes released from the different packaging cell lines were titrated by using NIH 3T3 cells or TE671 cells as target. Titers higher than  $1 \times 10^7$  lacZ i.u./ml were obtained for the best clones. Depending on the envelope specificities expressed in these cells, the new

TE671-based retroviral packaging cell lines were named TE-FLYE, TE-FLYA, TE-FLYRD, TE-FLY10A1, and TE-FLYGA and could express the MoMLV, MLV-4070A, RD114, MLV-10A1, and GALV env genes, respectively.

- 5 Assays for detecting replication-competent retroviruses (RCRs) were performed in the supernatants of these cells and were negative (less than 1/ml).

10 TE671 cells are very potent for transient expression resulting in more than 95% of cells expressing transgene three days after plasmid transfection (Hatzioannou and Cosset, unpublished data, (1996)). The ability of retroviral packaging cell lines to transiently produce retroviral vectors is of crucial importance for gene therapy where  
15 vectors carrying toxic gene have to be prepared. Transient expression of retroviral vectors was comparatively determined from cells of the TE-FLYA line and from the BING line (Pear et al., Proc Natl Acad Sci U S A 90, 8392-6 (1993)), a retroviral packaging cell line designed to  
20 transiently express retroviral vectors. Results (Table 8) showed that TE-FLYA cells were more efficient for transient expression of a lacZ retroviral vector hence resulting in higher titers.

25 **Table 8. Comparative study of transient production of lacZ vectors.**

packaging cell line	cell number <sup>a</sup>	% transfected cells <sup>b</sup>	transient titer <sup>c</sup>
BING	281	5.3	$2 \times 10^2$
TE-FLYA	117	35	$1.3 \times 10^3$

30 Cells were transfected by MFGnslacZ retroviral vectors with calcium phosphate precipitation method and titers of of lacZ vectors (c) released in cell

supernatant were determined as lacZ i.u./ml at day 3 following transfection. The relative number of cells (a) (average per microscope field) and the % of transfected cells (b) determined after X-gal staining are shown.

5 Retroviral vectors prepared from TE671-based packaging cell lines were analysed for their sensitivity to human-complement mediated inactivation. Experiments were conducted as previously described (Cosset et al., J. Virol. 69:7430-7436 (1995); see also Example 10 in this patent application)  
 10 by using three human sera of individual donors (Table 9). As expected MLV-A prepared from mouse 3T3 cells were highly sensitive to inactivation after 1 hr incubation with sera. In contrast, titers of lacZ vectors produced from TE-FLYRD  
 15 lacZ vectors from TE-FLYA cells were 1 to 30% of controls.

**Table 9. Human serum sensitivity of viruses produced from TE671-based packaging cell lines.**

Virus from:	hu56 <sup>a</sup>	hu57 <sup>a</sup>	BTS <sup>a</sup>
20 3T3/A	<0.2, <0.2	<0.2, <0.2	<0.2, <0.2
TE-FLYE	15, 7.8	16, 11	48, 60
TE-FLYA	1, 0.6	2.2, 7.1	28, 19
TE-FLYRD	17, 22	30, 44	54, 63

25 Three human fresh serum samples were tested in duplicate; hu56 (A+), hu57 (AB+), BTS (AB+). (a) % control (average for FCS and opti-MEM treatment) is shown.

CLAIMS:

1. A recombinant expression vector comprising a gene of interest and a selectable marker gene, wherein the selectable marker gene is arranged downstream of the gene of interest and a stop codon associated with the gene of interest is spaced from a start codon of said selectable marker gene at a distance which is sufficient to ensure that said selectable marker protein is expressed from the corresponding mRNA as a result of translation reinitiation.
2. A recombinant expression vector according to claim 1 wherein the vector is a viral vector.
3. A recombinant expression vector according to claim 2 wherein the vector is a retroviral vector.
4. A recombinant expression vector according to any one of claims 1 to 3 wherein the gene of interest is included as part of a viral packaging construct.
5. A recombinant expression vector according to any one of the preceding claims wherein the number of nucleotides in the space between the stop codon of the gene of interest and the start codon of the selectable marker is in the range of from 20 to 200 nucleotides.
6. A recombinant expression vector according to claim 5 wherein the number of nucleotides in the space between the stop codon of the gene of interest and the start codon of the selectable marker is in the range of from 60 to 80 nucleotides.
7. A process for producing a cell line in which a gene of interest is expressed, which process comprises:  
transforming host cells with an expression vector

according to any one of the claims 1 to 6; and  
selectable those cells where expression of the  
selection marker gene may be detected.

8. A process according to claim 7 wherein the host cell is a eukaryotic cell.
9. A host cell transformed with a recombinant expression vector according to any one of the claims 1 to 6.
10. A retroviral packaging cell line comprising a host cell transformed with a first and a second recombinant expression vector, said first recombinant expression vector having a packaging-deficient construct comprising a viral gag-pol gene and a first selectable marker gene downstream thereof, and said second recombinant expression vector having a packaging-deficient construct comprising a viral env gene and a second selectable marker gene downstream thereof; wherein the start codon of the first and second selectable markers are spaced from the stop codons of the viral gag-pol gene and the viral env gene respectively by a distance which ensures that said selectable marker protein is expressed from the corresponding mRNA as a result of translation reinitiation.
11. A retroviral packaging cell line according to claim 10 wherein the first selectable marker is a bsr selectable marker and the second selectable marker is a phleo selectable marker.
12. A retroviral packaging cell line according to any one of claims 10 or 11 wherein the packaging-deficient construct comprising the viral gag-pol gene and first selectable marker is the CeB (SEQ ID No 2) expression construct.

13. A retroviral packaging cell line according to any one of claims 10 or 11 wherein the packaging-deficient construct comprising the viral env gene and second selectable marker is the FBdelPASAF (SEQ ID No 5), the FBdelPMOSAF (SEQ ID No 6), the FBdelPGASAF (SEQ ID No 7), the FBdelPRDSAF (SEQ ID No 8), the FBdelPXSAF (Fig. 3), the FBdelP10A1SAF (Fig. 3), or the FBdelPVSVGSAF (Fig. 3) expression construct.
14. A retroviral packaging cell line according to any one of claims 10 or 11 wherein the recombinant expression vector is a packaging-deficient retroviral helper construct.
15. A retroviral packaging cell line according to claim 14 wherein the overlapping sequences between the genomes of the retroviral vector and the packaging-deficient construct is reduced by minimizing the extent of non-coding retroviral sequences in the packaging-deficient genome.
16. A retroviral packaging cell line according to any one of claims 10 to 15 wherein the viral gag-pol gene and the selectable marker are expressed under the control of a non-retroviral promoter.
17. A retroviral packaging cell line according to claim 16 wherein the promoter is fused to rabbit beta-1 globin intron.
18. A retroviral packaging cell line according to claim 16 or claim 17 wherein the promoter is a hCMV promoter.
19. A retroviral packaging cell line according to any one of claims 16 to claim 18 wherein the viral gag-pol gene and the selectable marker is a hCMV+intron (SEQ



ID No3) or a hCMV+intronkaSD (SEQ ID No 4) expression construct.

20. A retroviral packaging cell line according to anyone of claims 10 to 15 wherein the viral env gene and the selectable marker are under the control of a non-retroviral promoter.
21. A retroviral packaging cell line according to claim 20 wherein the promoter is fused to rabbit beta-1 globin intron.
22. A retroviral packaging cell line according to claim 20 or claim 21 wherein the promoter is a hCMV promoter.
23. A retroviral packaging cell line according any one of claims 20 to 22 wherein the viral env gene and the selectable marker is a CMV10A1 (SEQ ID No 9) expression construct.
24. A retroviral packaging cell line according to any one of claims 10 to 23 wherein the cell line is the HT1080 line, the TE671 line, the 3T3 line, the 293 line or the MV-1-LU line.
25. A retroviral packaging cell line according to anyone of claims 10 to 24 wherein the retroviral packaging cells comprises human HT1080 cells and express RD114 envelopes.
26. A retroviral packaging cell line according to anyone of claims 10 to 24 wherein the retroviral packaging cells comprises human TE671 cells and express RD114 envelopes.

27. A process for producing a retroviral packaging cell line in which a gene of interest is expressed, which process comprises:
- transforming host cells with a first and a second recombinant expression vector, said first recombinant expression vector having a packaging-deficient construct comprising a viral gag-pol gene and a first selectable marker gene downstream thereof, and said second recombinant expression vector having a packaging-deficient construct comprising a viral env gene and a second selectable marker gene downstream thereof; wherein the start codon of the first and second selectable markers are spaced from the stop codons of the viral gag-pol gene and the viral env gene respectively by a distance which ensures that said selectable marker protein is expressed from the corresponding mRNA as a result of translation reinitiation; and
- selecting transformed cells which express said first and/or second marker genes.
28. A packaging deficient construct for use in a process according to claim 27, which expresses a viral gag-pol gene and a selectable marker wherein a start codon of the selectable marker is spaced from a stop codon of the viral gag-pol gene by a distance which ensures that said selectable marker protein is expressed from the corresponding mRNA as a result of translation reinitiation.
29. A packaging deficient construct for use in a process according to claim 27, which expresses a viral env gene and a selectable marker gene; wherein a start codon of the selectable marker is spaced from a stop codon of the viral env gene by a distance which ensures that said selectable marker protein is

expressed from the corresponding mRNA as a result of translation reinitiation.

1/22

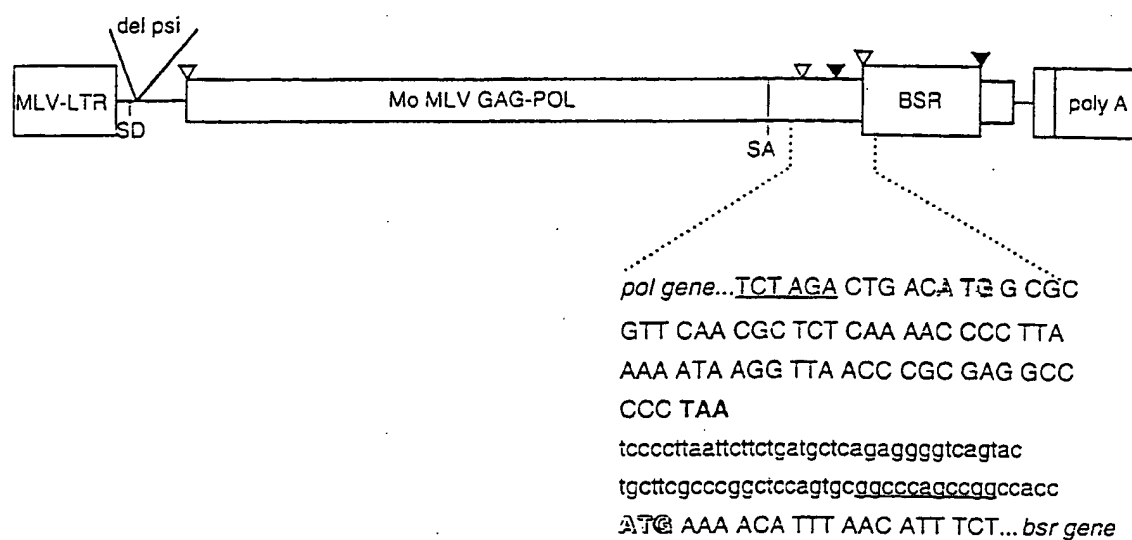


Figure 1. Schematic structure of CeB expression vector

2/22

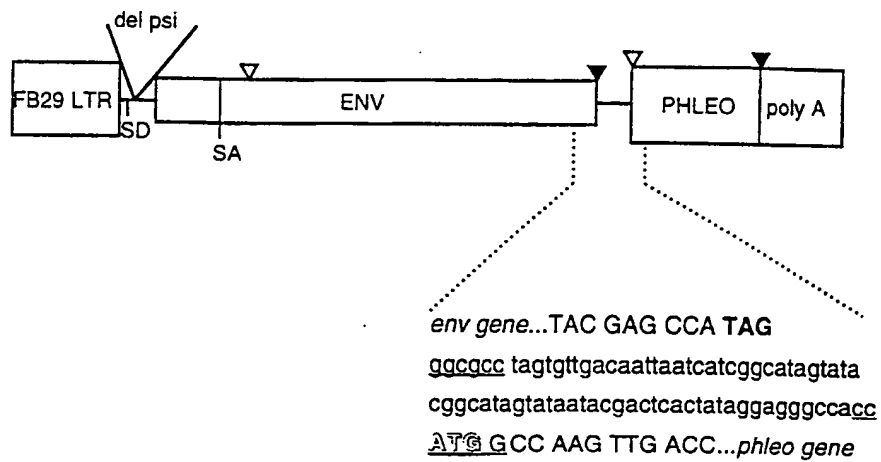


Figure 2. Schematic structure of FBdelPASF expression vector

3/22

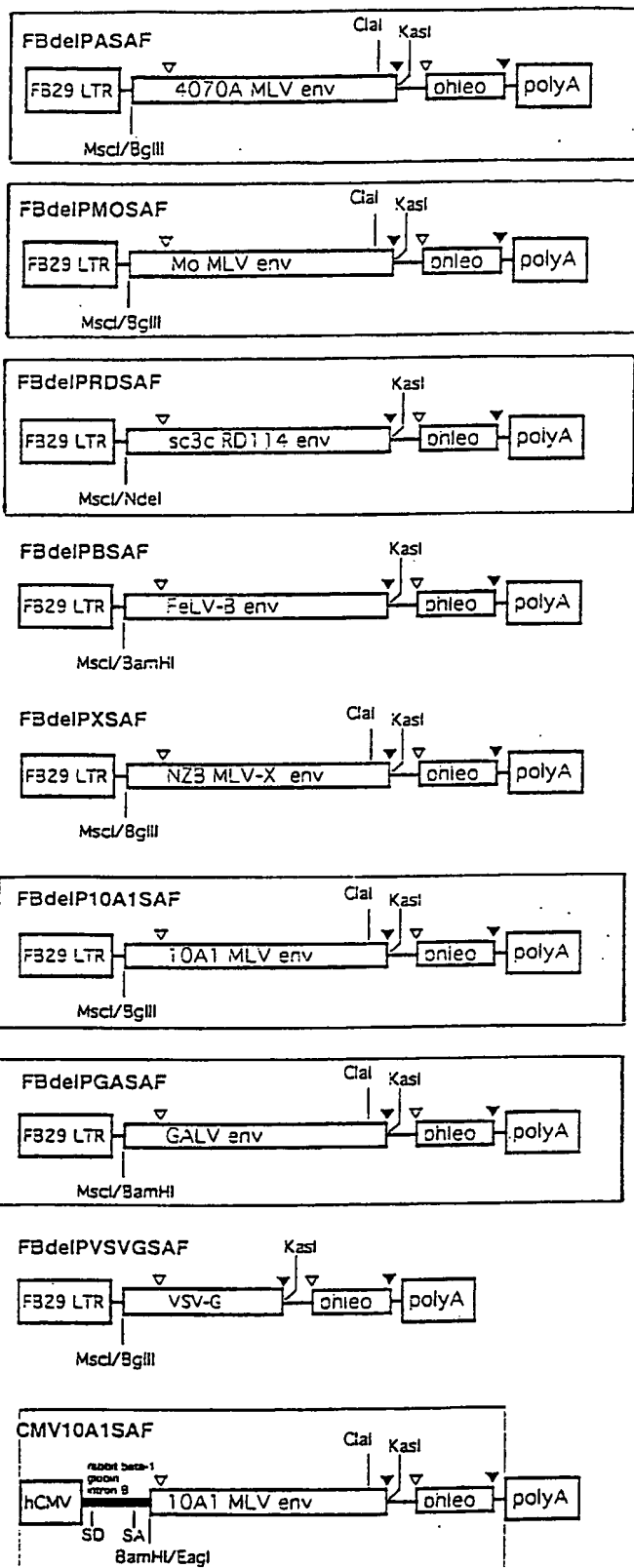


Figure 3. Schematic structure of env expression vectors

SUBSTITUTE SHEET (RULE 26)

4/22

NGAGCTCAGGACAGGTAGAAAAGATGAATAGAACAATAAAAGAGACCCTTACTAAATTGA 60  
CCTTAGAGACTGGCTTAAAAGATTGGAGACGCCTCCTATCTCTGGCTTTGTTAAGAGCCA 120  
GAAATACGCCCAACCGTTTTTCGGCTCACCCCATATGAAATCCTTTATGGGGGACCCCCC 180  
CTTTGTCAACCTTGCTCAATTCCTTCTCCCCCTCCGATCCTAAGACTGATTTACAAGCCC 240  
GACTAAAAGGGCTGCAAGGCGTGCAGGCCCAAATCTGGACACCCCTGGCCGAATTGTACC 300  
GGCCAGGACATCCACAAACTAGCCACCCATTTTCAGGTGGGAGACTCCGTGTACGTCCGGC 360  
GGCACCGCTCTCAAGGATTGGAGCCTCGTTGGAAGGGACCTTACATCGTCCTGCTGACCA 420  
CGCCACCGCCATAAAGGTTGACGGGATCGCCGCCTGGATTACGCATCGCACGCCAAGG 480  
CAGCCCCAAAAACCCCTGGACCAGAACTCCCAAAACCTGGAAGCTCCGCCGTTTCGGAGA 540  
ACCCTCTTAAGATAAGACTCTCCCGTGTCTGACTGCTAATCCACCTTGTCCTGTACTAA 600  
CCCAAAATGAACTCCCAACAGGAATGGTCATTTTATGTAGCCTAATAATAGTTCGGGCA 660  
GGGTTTGACGACCCCCGCAAGGCTATCGCATTAGTACAAAAACAACATGGTAAACCATGC 720  
GAATGCAGCGGAGGGCAGGTATCCGAGGCCCCACCGAACTCCATCCAACAGGTAACCTTGC 780  
CCAGGCAAGACGGCCTACTTAATGACCAACCAAAAATGGAAATGCAGAGTCACTCCAAAA 840  
ATCTCACCTAGCGGGGAGAACTCCAGAACTGCCCCCTGTAACTTTCCAGGACTCGATG 900  
CAGAGTTCTTGTATACTGAATACCGGCAATGCAGGCGAATTAATAAGACATACTACACG 960  
GCCACCTTGCTTAAAATACGGTCTGGGAGCCTCAACGAGGTACAGATATTACAAAACCCC 1020  
AATCAGCTCCTACAGTCCCCCTTGTAAGGGCTCTATAAATCAGCCCGTTTGCTGGAGTGCC 1080  
ACAGCCCCCATCCATATCTCCGATGGTGGAGGACCCCTCGATACTAAGAGAGTGTGGACA 1140  
GTCCAAAAAAGGCTAGAACAAATTCATAAGGCTATGACTCCTGAACTTCAATACCACCCC 1200  
TTAGCCCTGCCCCAAGTCAGAGATGACCTTAGCCTTGATGCACGGACTTTTGATATCCTG 1260  
AATACCACTTTTAGGTTACTCCAGATGTCCAATTTTAGCCTTGCCCCAAGATTGTTGGCTC 1320  
TGTTTAAAACTAGGTACCCCTACCCCTCTTGCGATACCCACTCCCTCTTTAACCTACTCC 1380  
CTAGCAGACTCCCTAGCGAATGCCTCCTGTCAGATTATACCTCCCTCTTGGTTCAACCG 1440  
ATGCAGTTCTCCAACCTCGTCCTGTTTATCTTCCCCTTTTATTAACGATACGGAACAAATA 1500  
GACTTAGGTGCAGTCACCTTTACTAACTGCACCTCTGTAGCCAATGTCAGTAGTCCTTTA 1560  
TGTGCCCTAAACGGGTCAGTCTTCCCTCTGTGGAAATAACATGGCATAACCTATTTACCC 1620  
CAAACTGGACCAGACTTTGCGTCCAAGCCTCCCTCCTCCCCGACATTGACATCAACCCG 1680  
GGGGATGAGCCAGTCCCCATTCTGCCATTGATCATTATATACATAGACCTAAACGAGCT 1740  
GTACAGTTCATCCCTTTACTAGCTGGACTGGGAATCACCGCAGCATTACACCACCGGAGCT 1800  
ACAGGCCTAGGTGTCTCCGTCACCCAGTATACAAAATTATCCCATCAGTTAATATCTGAT 1860  
GTCCAAGTCTTATCCGGTACCATAACAAGATTTACAAGACCAGGTAGACTCGTTAGCTGAA 1920  
GTAGTTCTCCAAAATAGGAGGGGACTGGACCTACTAACGGCAGAACAAAGGAGGAATTTGT 1980  
TTAGCCTTACAAGAAAAATGCTGTTTTTATGCTAACAAGTCAGGAATTGTGAGAAACAAA 2040  
ATAAGAACCCTACAAGAAGAATTACAAAAACGCAGGGAAAGCCTGGCAACCAACCCCTCTC 2100  
TGGACCGGGCTGCAGGGCTTTCTTCCGTACCTCCTACCTCTCCTGGGACCCCTACTCACC 2160  
CTCCTACTCATACTAACCATTGGGCCATGCGTTTTTCAGTCGCCTCATGGCCTTCATTAAT 2220  
GATAGACTTAATGTTGTACATGCCATGGTGTGCCCCAGCAATACCAAGCACTCAAAGCT 2280  
GAGGAAGAAGCTCAGGATTGAGCTTCCGGGACAAAAGCAGGGGGGAATGAGAAGTCAGAA 2340  
CCCCCACCTTTGCTACATAAATAACCGCTTTCATTTTCGCTTCTGTAAAACGCTTATGCG 2400  
CCCCACCTAGCCGGAAAGTCCCCAGCCGCTACGCAACCCGGGCCCCGAGTTGCATCAGC 2460  
CGTTCGCAACCCGGGCTCCGAGTTGCATCAGCCGAAAGAACTTCATTTCCCAAGCTT 2518

Fig.4

5/22

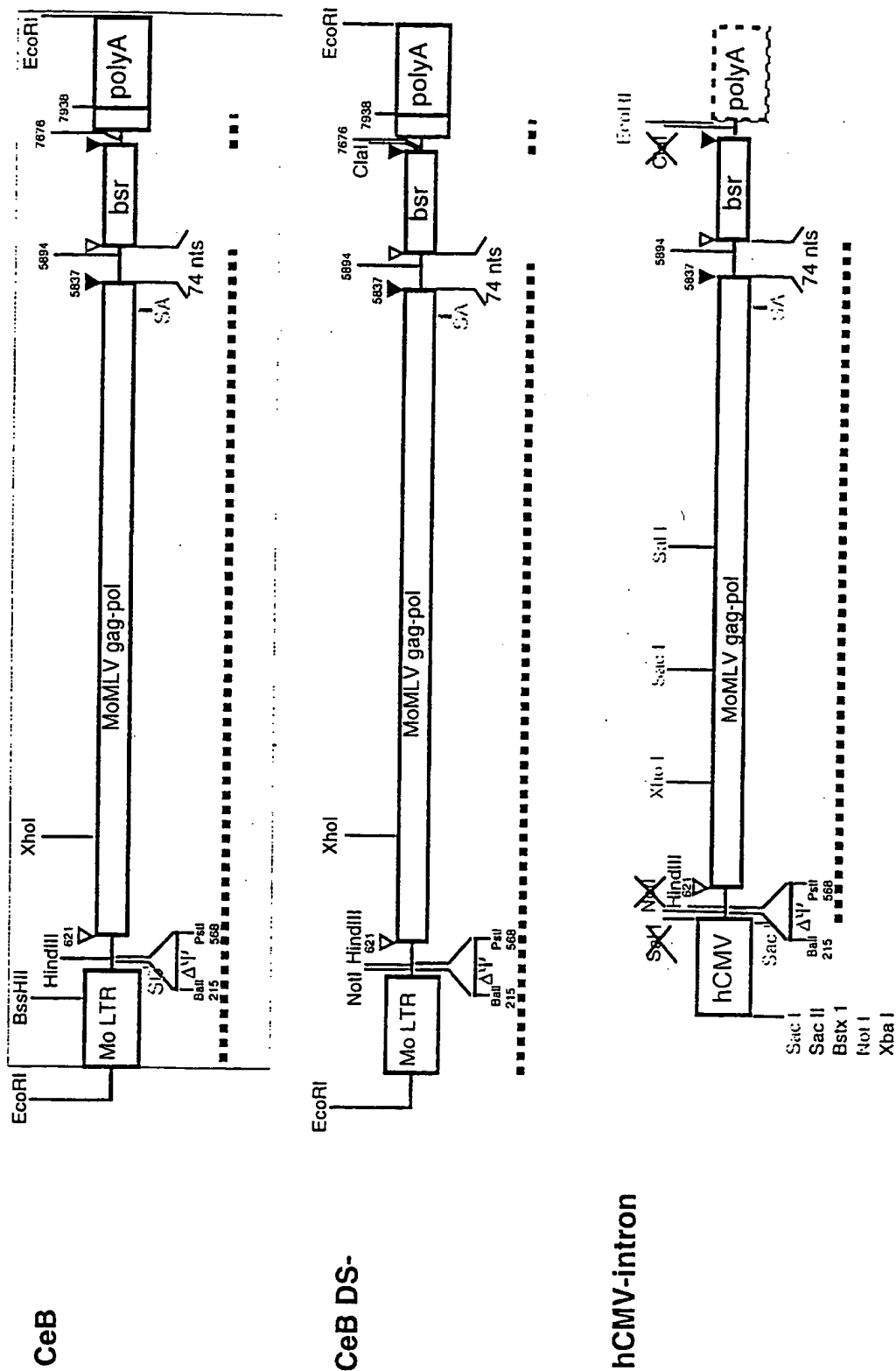


Figure 5. Genetic structure of gag-pol constructs (page 1/3)



6/22

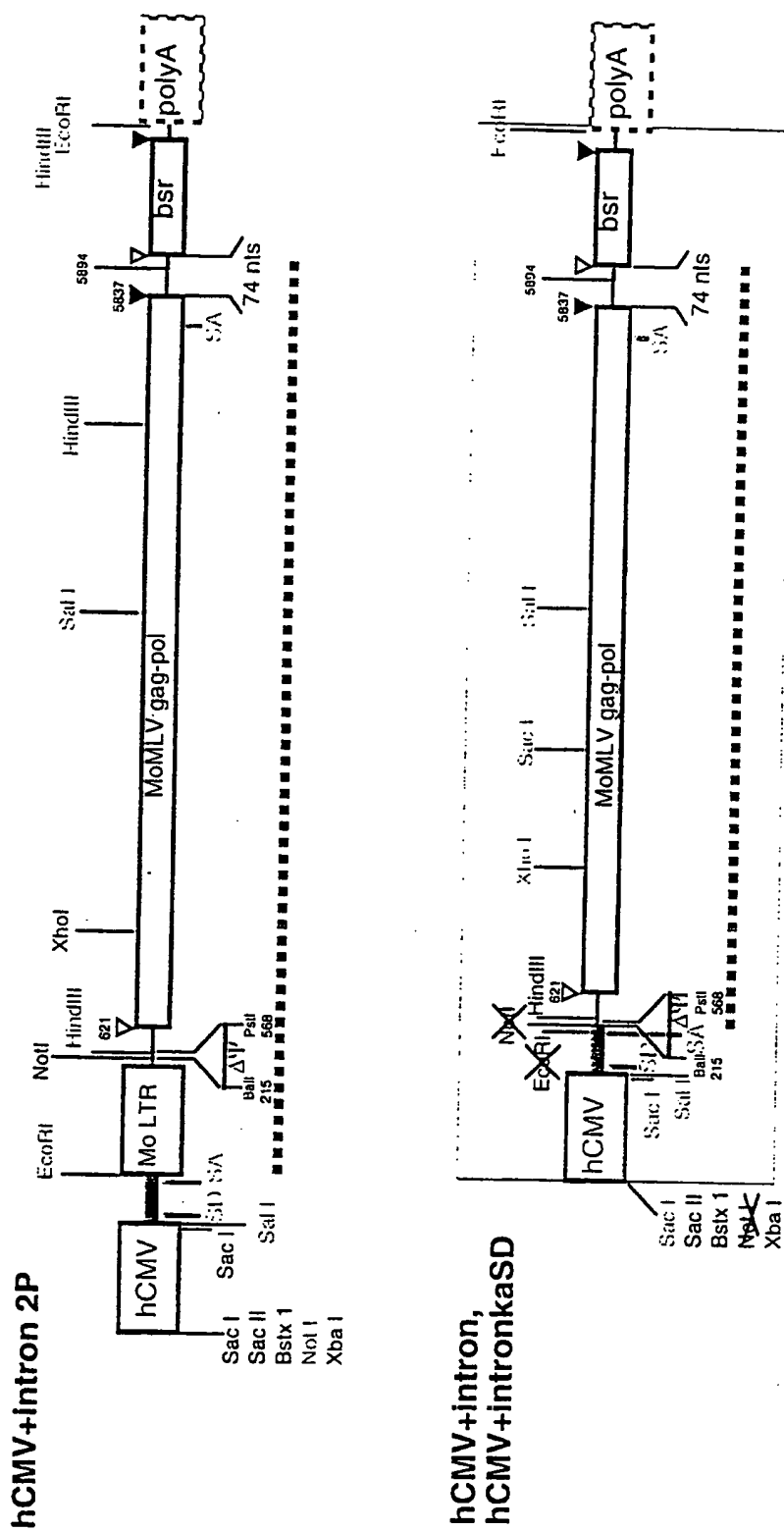


Figure 5. Genetic structure of gag-pol constructs (page 2/3)

7/22

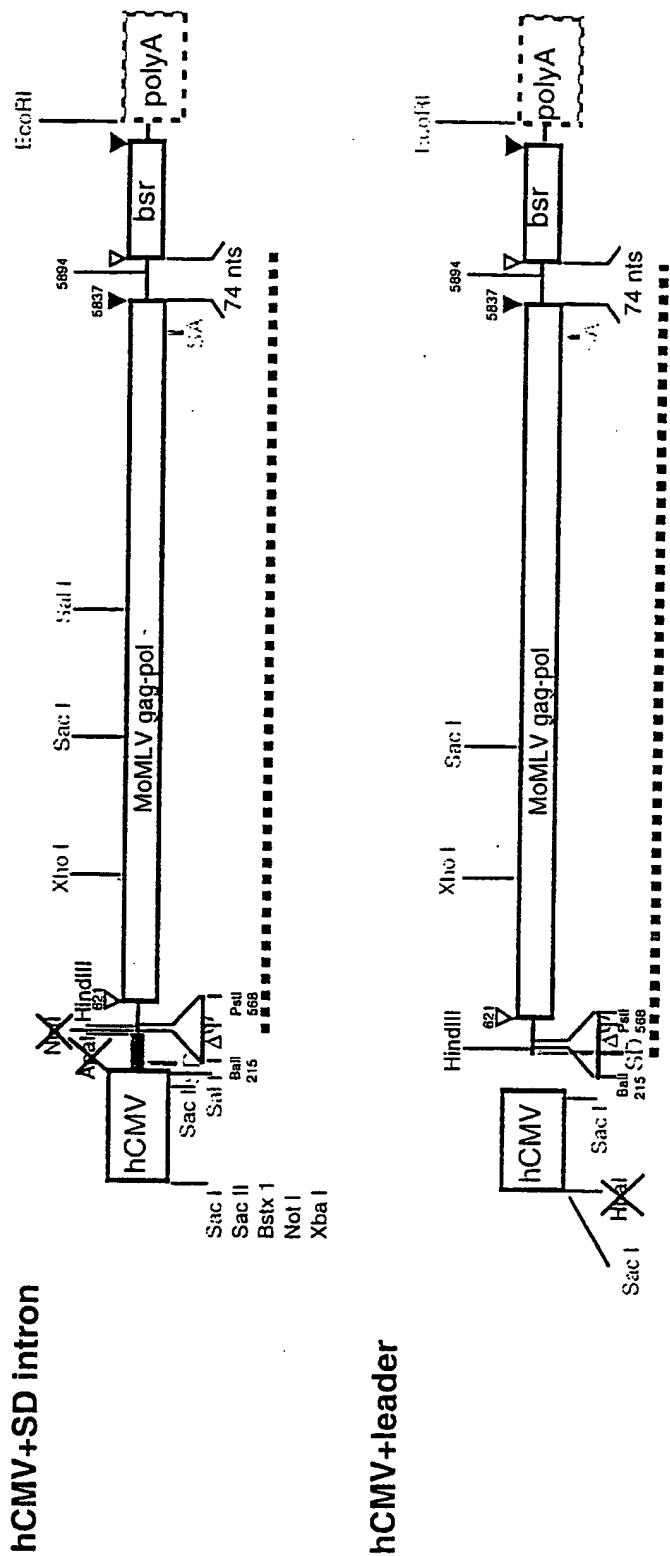


Figure 5. Genetic structure of gag-pol constructs (page 3/3)

Figure 6. CeB Sequence

8/22

1

AATGAAAGAC	CCCACCTGTA	GGTTTGGCAA	GCTAGCTTAA	GTAACGCCAT	TTTGCAAGGC	60
ATGGAAAAAT	ACATAACTGA	GAATAGAGAA	GTTTCAGATCA	AGGTCAGGAA	CAGATGGAAC	120
AGCTGAATAT	GGGCCAAACA	GGATATCTGT	GGTAAGCAGT	TCCTGCCCCG	GCTCAGGGCC	180
AAGAACAGAT	GGAACAGCTG	AATATGGGCC	AAACAGGATA	TCTGTGGTAA	GCAGTTCCTG	240
CCCCGGCTCA	GGGCCAAGAA	CAGATGGTCC	CCAGATGCGG	TCCAGCCCTC	AGCAGTTTCT	300
AGAGAACCAT	CAGATGTTTC	CAGGGTGCCC	CAAGGACCTG	AAATGACCCT	GTGCCCTATT	360
TGAACTAACC	AATCAGTTCG	CTTCTCGCTT	CTGTTTCGCG	GCTTCTGCTC	CCCGAGCTCA	420
ATAAAAGAGC	CCACAACCCC	TCACTCGGGG	CGCCAGTCCT	CCGATTGACT	GAGTCGCCCC	480
GGTACCCGTG	TATCCAATAA	ACCCTCTTGC	AGTTGCATCC	GACTTGTGGT	CTCGCTGTTC	540
CTTGGGAGGG	TCTCCTCTGA	GTGATTGACT	ACCCGTCAGC	GGGGGTCTTT	CATTTGGGGG	600
CTCGTCCGGG	ATTCGGGACG	CCCTGCCCAG	GGACCCACGA	CCCACCACCG	GGAGGTAAGC	660
TGGAAGCTTC	TGCAGCATCG	TTCTGTGTTG	TCTCTGTCTG	ACTGTGTTTC	TGTATTTGTC	720
TGAGAATATG	GGCCAGACTG	TTACCACTCC	CTTAAGTTTG	ACCTTAGGTC	ACTGGAAGA	780
TGTCGAGCGG	ATCGCTCACA	ACCAGTCGGT	AGATGTCAAG	AAGAGACGTT	GGGTTACCTT	840
CTGCTCCGCA	CATTGGCCAA	CCTTTAACGT	CGGATGGCCG	CGAGACGGCA	CCTTTAACCG	900
AGACCTCATC	ACCCAGGTTA	AGATCAAGGT	CTTTTCACCT	GGCCCGCATG	GACACCCAGA	960
CCAGGTCCCC	TACATCGTGA	CCTGGGAAGC	CTTGGCTTTT	GACCCCCCTC	CCTGGGTCAA	1020
GCCCTTTGTA	CACCCTAAGC	CTCCGCCCTC	TCTTCTCCCA	TCCGCCCCGT	CTCTCCCCCT	1080
TGAACCTCCT	CGTTGACCCC	CGCCTCGATC	CTCCCTTTAT	CCAGCCCTCA	CTCCTTCTCT	1140
AGGCGCCAAA	CCTAAACCTC	AAGTTCTTTC	TGACAGTGGG	GGGCCGCTCA	TCGACCTACT	1200
TACAGAAGAC	CCCCCGCCTT	ATAGGGACCC	AAGACCACCC	CCTTCGACA	GGGACGAAA	1260
TGGTGGAGAA	GCGACCCCTG	CGGGAGAGGC	ACCGGACCCC	TCCCCAATGG	CATCTCGCCT	1320
ACGTGGGAGG	CGGGAGCCCC	CTGTGGCCGA	CTCCACTACC	TCGCAGGCAT	TCCCCCTCCG	1380
CGCAGGAGGA	AACGGACAGC	TTCAATACTG	GCCGTTCTCC	TCTTCTGACC	TTTACAACCTG	1440
GAAAAATAAT	AACCTTCTTT	TTTCTGAAGA	TCCAGGTAAA	CTGACAGCTC	TGATCGAGTC	1500
TGTTCTCATC	ACCCATCAGC	CCACCTGGGA	CGACTGTCAG	CAGCTGTTGG	GGACTCTGCT	1560
GACCGGAGAA	GAAAAACAAC	GGGTGCTCTT	AGAGGCTAGA	AAGGCGGTGC	GGGGCGATGA	1620
TGGGCGCCCC	ACTCAACTGC	CCAATGAAGT	CGATGCGCCT	TTTCCCTCG	AGCGCCGAGA	1680
CTGGGATTAC	ACCACCCAGG	CAGGTAGGAA	CCACCTAGTC	CACATATCGC	AGTTGCTCCT	1740
AGCGGGTCTC	CAAAACGCGG	GCAGAAGCCC	CACCAATTTG	GCCAAAGTAA	AAGGAATAAC	1800
ACAAGGGCCC	AATGAGTCTC	CCTCGGCCTT	CCTAGAGAGA	CTTAAGGAAG	CCTATCGCAG	1860
GTACACTCCT	TATGACCCCTG	AGGACCCAGG	GCAAGAAAAC	AATGTGTCTA	TGCTTTTCAT	1920
TTGGCAGTCT	GCCCCAGACA	TTGGGAGAAA	GTTAGAGAGG	TTAGAAGATT	TAAAAACAA	1980
GACGCTTGGA	GATTTGTTTA	GAGAGGCAGA	AAAGATCTTT	AATAAACGAG	AAACCCCGGA	2040
AGAAAGAGAG	GAACGTATCA	GGAGAGAAAC	AGAGGAAAAA	GAAGAACGCC	GTAGGACAGA	2100
GGATGAGCAG	AAAGAGAAAG	AAAGAGATCG	TAGGAGACAT	AGAGAGATGA	GCAAGCTATT	2160
GGCCACTGTC	GTTAGTGGAC	AGAAACAGGA	TAGACAGGGA	GGAGAACGAA	GGAGGTCCCA	2220
ACTCGATCGC	GACCACTGTG	CCTACTGCAA	AGAAAAGGGG	CACCTGGGCTA	AAGATTGTCC	2280
CAAGAAACCA	CGAGGACCTC	GGGGACCAAG	ACCCAGACCC	TCCCTCCTGA	CCCTAGATGA	2340
CTAGGAGAGT	CAGGGTCAGG	AGCCCCCCCC	TGAACCCAGG	ATAACCCCTA	AAGTCGGGGG	2400
GCAACCCGTC	ACCTTCTCTG	TAGATACTGG	GGCCCAACAC	TCCGTGCTGA	CCCAAAATCC	2460
TGGACCCCTA	AGTGATAAGT	CTGCCTGGGT	CCAAGGGGCT	ACTGGAGGAA	AGCGGTATCG	2520
CTGGACCAAG	GATCGCAAAG	TACATCTAGC	TACCGGTAAG	GTCACCCACT	CTTTCCTCCA	2580
TGTACAGAGC	TGTCCTTATC	CTCTGTTAGG	AAGAGATTTG	CTGACTAAAC	TAAAAGCCCA	2640
AATCCACTTT	GAGGGATCAG	GAGCTCAGGT	TATGGGACCA	ATGGGGCAGC	CCCTGCAAGT	2700
GTTGACCCCTA	AATATAGAAG	ATGAGCATCG	GCTACATGAG	ACCTCAAAAG	AGCCAGATGT	2760
TTCTCTAGGG	TCCACTGGG	TGCTCTGATT	TCCTCAGGCC	TGGGCGGAAA	CCGGGGGCGT	2820
GGGACTGGCA	GTTTCGCCAAG	CTCCTCTGAT	CATACCTCTG	AAAGCAACCT	CTACCCCCGT	2880
GTCCATAAAA	CAATACCCCA	TGTCACAAGA	AGCCAGACTG	GGGATCAAGC	CCCACATACA	2940
GAGACTGTTG	GACCAGGGAA	TACTGGTACC	CTGCCAGTCC	CCCTGGAACA	CGCCCTTGCT	3000
ACCCGTTAAG	AAACCAGGGA	CTAATGATTA	TAGGCCTGTC	CAGGATCTGA	GAGAAGTCAA	3060
CAAGCGGGTG	GAAGACATCC	ACCCACCCGT	GCCCCAACCT	TACAACCTCT	TGAGCGGGCT	3120
CCCACCGTCC	CACCACTGGT	ACACTGTGCT	TGATTTAAAG	GATGCCTTTT	TCTGCCTGAG	3180
ACTCCACCCC	ACCAGTCAGC	CTCTCTTCGC	CTTTGAGTGG	AGAGATCCAG	AGATGGGAAT	3240
CTCAGGACAA	TTGACCTGGA	CCAGACTCCC	ACAGGTTTTC	AAAAACAGTC	CCACCCTGTT	3300
TGATGAGGCA	CTGCACAGAG	ACCTAGCAGA	CTTCCGATC	CAGCACCCAG	ACTTGATCCT	3360
GCTACAGTAC	GTGGATGACT	TACTGCTGGC	CGCCACTTCT	GAGCTAGACT	GCCAAACAAG	3420
TACTCGGGCC	CTGTTACAAA	CCCTAGGGAA	CCTCGGGTAT	CGGGCCTCGG	CCAAGAAAGC	3480
CCAAATTTCG	CAGAAACAGG	TCAAGTATCT	GGGGTATCTT	CTAAAAGAGG	GTCAGAGATG	3540
GCTGACTGAG	GCAGAAAAAG	AGACTGTGAT	GGGGGACCTT	ACTCGGAAGA	CCCCTCGACA	3600
ACTAAGGGAG	TTCTTAGGGA	CGGCAGGCTT	CTGTCGCCCT	TGGATCCCTG	GGTTTGCAGA	3660
AATGGCAGCC	CCCTTGTACC	CTCTCACCAA	AACGGGGACT	CTGTTTAATT	GGGGCCGAGA	3720
CCAACAAAAG	GCCTATCAAG	AAATCAAGCA	AGCTCTTCTA	ACTGCCCCAG	CCCTGGGGTT	3780
GCCAGATTTG	ACTAAGCCCT	TTGAACCTTT	TGTCAGCAGG	AAGCAGGGCT	ACGCCAAAGG	3840
TGTCCTAACG	CAAAAACCTG	GACCTTGGCG	TCGGCCGGTG	GCCTACCTGT	CCAAAAGCT	3900
AGACCCAGTA	GCAGCTGGGT	GGCCCCCTTG	CCTACCGATG	GTAGCAGCCA	TTGCCGTACT	3960
GACAAAGGAT	GCAGGCAAGC	TAACCATGGG	ACAGCCACTA	GTCATTCTGG	CCCCCATGCT	4020
AGTAGAGGCA	CTAGTCAAAAC	AACCCCCCGA	CCGCTGGCTT	TCCAACGCCC	GGATGACTCA	4080

Figure 6. CeB Sequence

9/22

2

CTATCAGGCC	TTGCTTTTGG	ACACGGACCG	GGTCCAGTTC	GGACCGGTGG	TAGCCCTGAA	4140
CCCGGCTACG	CTGCTCCCAC	TGCCTGAGGA	AGGGCTGCAA	CACAACTGCC	TTGATATCCT	4200
GGCCGAAGCC	CACGGAACCC	GACCCGACCT	AACGGACCAG	CCGCTCCCAG	ACGCCGACCA	4260
CACCTGGTAC	ACGGATGGAA	GCAGTCTCTT	ACAAGAGGGA	CAGCGTAAGG	CGGGAGCTGC	4320
GGTGACCACC	GAGACCGAGG	TAATCTGGGC	TAAAGCCCTG	CCAGCCGGGA	CATCCGCTCA	4380
GCGGGCTGAA	CTGATAGCAC	TCACCCAGGC	CCTAAAGATG	GCAGAAGGTA	AGAAGCTAAA	4440
TGTTTATACT	GATAGCCGTT	ATGCTTTTGC	TACTGCCCAT	ATCCATGGAG	AAATATACAG	4500
AAGGCGTGGG	TTGCTCACAT	CAGAAGGCAA	AGAGATCAAA	AATAAAGACG	AGATCTTGGC	4560
CCTACTAAAA	GCCCTCTTTC	TGCCCAAAAG	ACTTAGCATA	ATCCATTGTC	CAGGACATCA	4620
AAAGGGACAC	AGGCCCGAGG	CTAGAGGCAA	CCGGATGGCT	GACCAAGCGG	CCCGAAAGGC	4680
AGCCATCACA	GAGACTCCAG	ACACCTCTAC	CCTCCTCATA	GAAAATTTCAT	CACCCCTACAC	4740
CTCAGAACAT	TTTCATTACA	CAGTGACTGA	TATAAAGGAC	CTAACCAAGT	TGGGGGCCAT	4800
TTATGATAAA	ATCAAGAAGT	ATTGGGTCTA	CCAAGGAAAA	CCTGTGATGC	CTGACCAGTT	4860
TACTTTTGAA	TTATTAGACT	TTCTTCATCA	GCTGACTCAC	CTCAGCTTCT	CAAAAATGAA	4920
GGCTCTCCTA	GAGAGAAGCC	ACAGTCCCTA	CTACATGCTG	AACCGGGATC	GAACACTCAA	4980
AAATATCACT	GAGACCTGCA	AAGCTTGTGC	ACAAGTCAAC	GCCAGCAAGT	CTGCCGTTAA	5040
ACAGGCAACT	AGGGTCCGCG	GGCATCGGCC	CGGCACCTCAT	TGGGAGATCG	ATTTACCCGA	5100
GATAAAGCCC	GGATTGTATG	GCTATAAATA	TCTTCTAGTT	TTTATAGATA	CCTTTTCTGG	5160
CTGGATAGAA	GCCTTCCCAA	CCAAGAAAGA	AACCGCCAAG	GTCGTAACCA	AGAAGCTACT	5220
AGAGGAGATC	TTCCCCAGGT	TCGGCATGCC	TCAGGTATTG	GGAAGTGACA	ATGGGCCTGC	5280
CTTCGTCTCC	AAGGTGAGTC	AGACAGTGGC	CGACTGTGTG	GGGATTGATT	GGAAATTACA	5340
TTGTGCATAC	AGACCCCAAA	GCTCAGGCCA	GGTAGAAAGA	ATGAATAGAA	CCATCAAGGA	5400
GACTTTAACT	AAATTAACGC	TTGCAACTGG	CTCTAGAGAC	TGGGTGCTCC	TACTCCCCTT	5460
AGCCCTGTAC	CGAGCCCGCA	ACACGCCGGG	CCCCATGGC	CTCACCCCAT	ATGAGATCTT	5520
ATATGGGGCA	CGCCCGCCCC	TTGTAAACTT	CCCTGACCCCT	GACATGACAA	GAGTTACTAA	5580
CAGCCCCCTCT	CTCCAAGCTC	ACTTACAGGC	TCTCTACTTA	GTCCAGCACG	AAGTCTGGAG	5640
ACCTCTGGCG	GCAGCCTACC	AAGAACAAC	GGACCGACCG	GTGGTACCTC	ACCCTTACCG	5700
AGTCGGCGAC	ACAGTGTGGG	TCCGCCGACA	CCAGACTAAG	AACCTAGAAC	CTCGCTGGAA	5760
AGGACCTTAC	ACAGTCTGTC	TGACCACCCC	CACCGCCCTC	AAAGTAGACG	GCATCGCAGC	5820
TTGGATACAC	GCCGCCACG	TGAAGGCTGC	CGACCCCGGG	GGTGGACCAT	CCTCTAGACT	5880
GACATGGCGC	GTTCAACGCT	CTCAAAACCC	CTTAAAAATA	AGGTTAACCC	GCGAGGCCCC	5940
CTAATCCCTT	TAATTCTTCT	GATGCTCAGA	GGGGTCAGTA	CTGCTTCGCC	CGGCTCCAGT	6000
GCGGCCGACG	GAAACACATT	AACATTTCTC	AACAAGATCT	AGAATTAGTA		6060
GAAGTAGCGA	CAGAGAAGAT	TACAATGCTT	TATGAGGATA	ATAAACATCA	TGTGGGAGCG	6120
GCAATTCTGT	CGAAAACAGG	AGAAATCATT	TCGGCAGTAC	ATATTGAAGC	GTATATAGGA	6180
CGAGTAACGT	TTTGTGCAGA	AGCCATTGCG	ATTGGTAGTG	CAGTTTCGAA	TGGACAAAAG	6240
GATTTTGACA	CGATTGTAGC	TGTTAGACAC	CCTTATTCTG	ACGAAGTAGA	TAGAAGTATT	6300
CGAGTGGTAA	GTCCTTGTGG	TATGTGTAGG	GAGTTGATTT	CAGACTATGC	ACCAGATTGT	6360
TTTGTGTTAA	TAGAAATGAA	TGGCAAGTTA	GTCAAAACTA	CGATTGAAGA	ACTCATTCCA	6420
CTCAATATA	CCCGAAATTA	AAAGTTTTC	CACCAAGCTT	ATCGATTAGT	CCAATTTGTT	6480
AAAGACAGGA	TATCAGTGGT	CCAGGCTCTA	GTTTTGACTC	AACAATATCA	CCAGCTGAAG	6540
CCTATAGAGT	ACGAGCCATA	GATAAAATAA	AAGATTTTAT	TTAGTCTCCA	GAAAAAGGGG	6600
GGAATGAAAG	ACCCACCTG	TAGGTTTGGC	AAGCTAGCTT	AAGTAACGCC	ATTTTGAAG	6660
GCATGGAAAA	ATACATAACT	GAGAATAGAG	AAGTTCAGAT	CAAGGTCAGG	AACAGATGGA	6720
ACAGTCGAGA	ACTTGTTTAT	TGCAGCTTAT	AATGGTTACA	AATAAAGCAA	TAGCATCACA	6780
AATTTACACA	ATAAAGCATT	TTTTTCACTG	CATTCTAGTT	GTGGTTTGTC	CAAACTCATC	6840
AATGTATCTT	ATCATGTCTG	GATCCCCAGG	AAGCTCCTCT	GTGTCTTCAT	AAACCCTAAC	6900
CTCCTCTACT	TGAGAGGACA	TTCCAATCAT	AGGCTGCCCA	TCCACCTCT	GTGTCTCTCT	6960
GTTAATTAGG	TCACCTTAACA	AAAAGGAAAT	TGGGTAGGGG	TTTTTCACAG	ACCGCTTTCT	7020
AAGGGTAATT	TTAAATATC	TGGGAAGTCC	CTTCCACTGC	TGTGTTCCAG	AAGTGTGGT	7080
AAACAGCCCA	CAAATGTCAA	CAGCAGAAAC	ATACAAGCTG	TCAGCTTTGC	ACAAGGGCCC	7140
AACACCCCTGC	TCATCAAGAA	GCACTGTGGT	TGCTGTGTTA	GTAATGTGCA	AAACAGGAGG	7200
CACATTTTCC	CCACCTGTGT	AGGTTCCAAA	ATATCTAGTG	TTTTCATTTT	TACTTGGATC	7260
AGGAACCCAG	CACCTCCACTG	GATAAGCATT	ATCCTTATCC	AAAACAGCCT	TGTGGTCAGT	7320
GTTTATCTGC	TGACTGTCAA	CTGTAGCATT	TTTTGGGGTT	ACAGTTTGAG	CAGGATATTT	7380
GGTCTGTAG	TTTGCTAACA	CACCCTGCAG	CTCCAAAGTT	TCCCCACCAA	CAGCAAAAAA	7440
ATGAAAAATTT	GACCCCTGAA	TGGGTTTTCC	AGCACCATT	TCATGAGTTT	TTTGTGTCCC	7500
TGAATGCAAG	TTTAACATAG	CAGTTACCCC	AATAACCTCA	GTTTTAACAG	TAACACCTTC	7560
CCACATCAAA	ATATTTCCAC	AGGTTAAGTC	CTCATTTAAA	TTAGGCAAG	GAATTC	7616

Figure 7. hCMV+intron Sequence

1

AGATCTCCCG	ATCCCCCTATG	GTCGACTCTC	AGTACAATCT	GCTCTGATGC	CGCATAGTTA	60
AGCCAGTATC	TGCTCCCTGC	TTGTGTGTTG	GAGGTGCGTG	AGTAGTGCGC	GAGCAAAATT	120
TAAGCTACAA	CAAGGCAAGG	CTTGACCGAC	AATTGCATGA	AGAATCTGCT	TAGGGTTAGG	180
CGTTTTGCGC	TGCTTCGCGA	TGTACGGGCC	AGATATACGC	GTGACATTG	ATTATTGACT	240
AGTTATTAAT	AGTAATCAAT	TACGGGGTCA	TTAGTTCATA	GCCCATATAT	GGAGTTCCGC	300
GTTACATAAC	TTACGGTAAA	TGCCCCGCTT	GGCTGACCGC	CCAACGACCC	CCGCCCATTG	360
ACGTCAATAA	TGACGTATGT	TCCCATAGTA	ACGCCAATAG	GGACTTTCCA	TTGACGTCAA	420
TGGGTGGACT	ATTTACGGTA	AACTGCCCC	TTGGCAGTAC	ATCAAGTGTA	TCATATGCCA	480
AGTACGCCCC	CTATTGACGT	CAATGACGGT	AAATGGCCCC	CCTGGCATT	TGCCCAGTAC	540
ATGACCTTAT	GGGACTTTCC	TACTTGCCAG	TACATCTACG	TATTAGTCAT	CGCTATTACC	600
ATGGTGATGC	GGTTTTGGCA	GTACATCAAT	GGCGTGGAT	AGCGGTTTGA	CTCACGGGGA	660
TTTCCAAGTC	TCCACCCCAT	TGACGTCAAT	GGGAGTTTGT	TTTGGCACCA	AAATCAACCG	720
GACTTTCCAA	AATGTCGTAA	CAACTCCGCC	CCATTGACGC	AAATGGGCGG	TAGGCGTGTA	780
CGGTGGGAG	TCTATATAAG	CAGAGCTCTC	TGGCTAACTA	GAGAACCAC	TGCTTAACTG	840
GCTTATCGAA	ATCTCGACTG	AGAAGTTTCT	GGTGAGTTTG	GGGACCCTTG	ATTGTTCTTT	900
CTTTTTTCGCT	ATTGTAAAAT	TCATGTTATA	TGGAGGGGGC	AAAGTTTTC	GGGTGTTGTT	960
TAGAATGGGA	AGATGTCCCT	TGTATCACCA	TGGACCCTCA	TGATAATTTT	GTTTCTTTCA	1020
CTTTCTACTC	TGTTGACAA	CATTGCTCTC	TCTTATTTTC	TTTTCATTTT	CTGTAACCTT	1080
TTCGTTAAAC	TTTAGCTTGC	ATTTGTAACG	AATTTTAA	TTCACTTTTG	TTTATTTGTC	1140
AGATTGTAAG	TACTTTCTCT	AATCACTTTT	TTTTCAAGGC	AATCAGGGTA	TATTATATTG	1200
TACTTCAGCA	CAGTTTTAGA	GAACAATTGT	TATAATTAAA	TGATAAGGTA	GAATATTTCT	1260
GCATATAAAT	TCTGGCTGGC	GTGGAAATAT	TCTTATTTGGT	AGAAACAAC	ACATCCTGGT	1320
CATCATCTCG	ACCTTCTCTT	TATGGTTACA	ATGATATACA	CTGTTTGAGA	TGAGGATAAA	1380
ATACTCTGAG	TCCAAACCGG	GCCCCCTCTC	TAACCATGTT	CATGCCTTCT	TCTTTTCTCT	1440
ACAGCTCCTG	GGCAACGTGC	TGGTTGTTGT	GCTGTCTCAT	CATTTTGGCA	AGAATTGGCC	1500
GCAAGCTTCT	GCAGCATCGT	TCTGTGTTGT	CTCTGTCTGA	CTGTGTTTCT	GTATTGTGCT	1560
GAGAATATGG	GCCAGACTGT	TACCACCTCC	TTAAGTTTGA	CCTTAGGTCA	CTGGAAAGAT	1620
GTCGAGCGGA	TCGCTCACAA	CCAGTCGGTA	GATGTCAAAG	AGAGACGTTG	GTTTACCTTC	1680
TGCTCTGCAG	AATGGCCAAC	CTTTAACGTC	GGATGGCCGC	GAGACGGCAC	CTTTAACCGA	1740
GACCTCATCA	CCCAGGTTAA	GATCAAGGTC	TTTTACCTTG	GCCCCGATGG	ACACCCAGAC	1800
CAGGTCCCTT	ACATCGTGAC	CTGGGAAGCC	TTGGCTTTTG	ACCCCCCTCC	CTGGGTCAAG	1860
CCCTTTGTAC	ACCCTAAGCC	TCCGCCCTCT	CTTCTTCCAT	CCGCCCCGTC	TCTCCCTCTT	1920
GAACCTCCTC	GTTCGACCCC	GCCTCGATCC	TCCCTTTATC	CAGCCCTCAC	TCCTTCTCTA	1980
GGCGCCAAAC	CTAAACCTCA	AGTTCTTTCT	GACAGTGGGG	GGCCGCTCAT	CGACCTACTT	2040
ACAGAAAGACC	CCCCGCCTTA	TAGGGACCCA	AGACACCCCC	CTTCCGACAG	GGACGGAAAT	2100
GGTGGAGAAG	CGACCCCTGC	GGGAGAGGCA	CCGGACCCCT	CCCCAATGGC	ATCTCGCCTA	2160
CGTGGGAGAC	GGGAGCCCCC	TGTGGCCGAC	TCCACTACCT	CGCAGGCATT	CCCCCTCCGC	2220
GCAGGAGGAA	ACGGACAGCT	TCAATACTGG	CCGTCTCTCT	CTTCTGACCT	TTACAACCTG	2280
AAAAATAATA	ACCTTCTTTT	TTCTGAAGAT	CCAGGTAAAC	TGACAGCTCT	GATCGAGTCT	2340
GTTCTCATCA	CCCATCAGCC	CACTTGGGAC	GACTCTCAGC	AGCTGTTGGG	GACTCTGCTG	2400
ACCGGAGAAG	AAAAACAACG	GGTGCTCTTA	GAGGTAGAA	AGGCGGTGCG	GGGCGATGAT	2460
GGGCGCCCCA	CTCAACTGCC	CAATGAAGTC	GATGCCGCTT	TTCCCTTCGA	GCGCCAGAC	2520
TGGGATTACA	CCACCCAGGC	AGGTAGGAAC	CACCTAGTCC	ACTATCGCCA	GTTGCTCCTA	2580
GCGGGTCTCC	AAAACGCGGG	CAGAAGCCCC	ACCAATTTGG	CCAAGGTAAA	AGGAATAACA	2640
CAAGGGCCCC	ATGAGTCTCC	CTCGGCCTTC	CTAGAGAGAC	TTAAGGAAGC	CTATCGCAGG	2700
TACACTCCTT	ATGACCCTGA	GGACCCAGGG	CAAGAACTA	ATGTGTCTAT	GTCTTTCATT	2760
TGGCAGTCTG	CCCCAGACAT	TGGGAGAAAG	TTAGAGAGGT	TAGAAGATTT	AAAAACAAG	2820
ACGCTTGGAG	ATTTGGTTAG	AGAGGCAGAA	AAGATCTTTA	ATAAACGAGA	AACCCCGGAA	2880
GAAAGAGAGG	AACGTATCAG	GAGAGAAACA	GAGGAAAAAG	AAGAACGCCG	TAGGACAGAG	2940
GATGAGCAGA	AAGAGAAAGA	AAGAGATCGT	AGGAGACATA	GAGAGATGAG	CAAGCTATTG	3000
GCCACTGTCG	TTAGTGGACA	GAAACAGGAT	AGACAGGGAG	GAGAACGAAG	GAGGTCCCAA	3060
CTCGATCGCG	ACCAGTGTGC	CTACTGCAAA	GAAAAGGGGC	ACTGGGCTAA	AGATTGTCCC	3120
AAGAAACCAC	GAGGACCTCG	GGGACCAAGA	CCCCAGACCT	CCCTCCTGAC	CCTAGATGAC	3180
TAGGGAGGTC	AGGGTCAGGA	GCCCCCCCC	GAACCCAGGA	TAACCCCTCA	AGTCGGGGGG	3240
CAACCCGTC	CCTTCTGGT	AGATACTGGG	GCCCAACACT	CCGTGCTGAC	CCAAAATCCT	3300
GGACCCCTAA	GTGATAAGTC	TGCCGTTGGT	CAAGGGGCTA	CTGGAGGAAA	GCGGTATCGC	3360
TGGACCACGG	ATCCGAAAGT	ACATCTAGCT	ACCGGTAAGG	TCACCCACTC	TTTCTCCAT	3420
GTACCAGACT	GTCCCTATCC	TCTGTTAGGA	AGAGATTTGC	TGACTAAACT	AAAAGCCCAA	3480
ATCCACTTTG	AGGGATCAGG	AGCTCAGGTT	ATGGGACCAA	TGGGGCAGCC	CCTGCAAGTG	3540
TTGACCCATA	ATATAGAAGA	TGAGCATCGG	CTACATGAGA	CCTCAAAAGA	GCCAGATGTT	3600
TCTCTAGGGT	CCCATGGCT	GTCTGATTTT	CCTCAGGCCT	GGGCGGAAAC	CGGGGGCATG	3660
GGACTGGCAG	TTCCGCAAGC	TCTCTGATC	ATACCTCTGA	AAGCAACCTC	TACCCCGGTG	3720
TCCATAAAAC	AATACCCCAT	GTCAACAAGA	GCCAGACTGG	GGATCAAGCC	CCACATACAG	3780
AGACTGTTGG	ACCAGGGAAT	ACTGTTACCC	TGCCAGTCCC	CCTGGAACAC	GCCCCGTGTA	3840
CCCCTTAAGA	AACCAGGGAC	TAATGATTAT	AGGCCTGTCC	AGGATCTGAG	AGAAGTCAAC	3900
AAGCGGGTGG	AAGACATCCA	CCCCACCGTG	CCCAACCTTT	ACAACCTCTT	GAGCGGGCTC	3960
CCACCGTCCC	ACCAGTGGTA	CACTGTGCTT	GATTTAAAGG	ATGCCTTTTT	CTGCCTGAGA	4020
CTCCACCCCA	CCAGTCAGCC	TCTCTTCGCC	TTTGAGTGGA	GAGATCCAGA	GATGGGAATC	4080

11/22

2

Figure 7. hCMV+intron Sequence

TCAGGACAAT	TGACCTGGAC	CAGACTCCCA	CAGGGTTTCA	AAAACAGTCC	CACCCTGTTT	4140
GATGAGGCAC	TGCACAGAGA	CCTAGCAGAC	TTCCGGATCC	AGCACCCAGA	CTTGATCCTG	4200
CTACAGTACG	TGGATGACTT	ACTGCTGGCC	GCCACTTCTG	AGCTAGACTG	CCAACAAGGT	4260
ACTCGGGCCC	TGTTACAAAC	CCTAGGGAAC	CTCGGGTATC	GGGCCTCGGC	CAAGAAAGCC	4320
CAAATTTGCC	AGAAACAGGT	CAAGTATCTG	GGGTATCTTC	TAAAAGAGGG	TCAGAGATGG	4380
CTGACTGAGG	CCAGAAAAGA	GACTGTGATG	GGGCAGCCTA	CTCCGAAGAC	CCCTCGACAA	4440
CTAAGGGAGT	TCCTAGGGAC	GGCAGGCTTC	TGTCGCCTCT	GGATCCCTGG	GTTTGCAGAA	4500
ATGGCAGCCC	CCTTGTACCC	TCTACCAAAA	ACGGGGACTC	TGTTTAATTG	GGGCCAGAC	4560
CAACAAAAGG	CCTATCAAGA	AATCAAGCAA	GCTCTTCTAA	CTGCCCCAGC	CCTGGGGTTG	4620
CCAGATTTGA	CTAAGCCCTT	TGAACTCTTT	GTCGACGAGA	AGCAGGGCTA	CGCCAAAGGT	4680
GTCCTAACGC	AAAAACTGGG	ACCTTGGCGT	CGGCCGGTGG	CCTACCTGTC	CAAAAAGCTA	4740
GACCCAGTAG	CAGACTGGGTG	GCCCCCTTGC	GCGCTGCAAC	TAGCAGCCAT	TGCCGTACTG	4800
ACAAAGGATG	CAGGCAAGCT	AACCATGGGA	CAGCCACTAG	TCATTCTGGC	CCCCATGCA	4860
GTAGAGGCAC	TAGTCAAACA	ACCCCCCGAC	CGCTGGCTTT	CCAACGCCCG	GATGACTCAC	4920
TATCAGGCCCT	TGCTTTTGGA	CACGGACCGG	GTCCAGTTCC	GACCGGTGGT	AGCCCTGAAC	4980
CCGGCTACGC	TGCTCCCACT	GCCTGAGGAA	GGCTGCTAAC	ACAACTGCCT	TGATATCCTG	5040
GCCGAAGCCC	ACGGAACCCG	ACCCGACCTA	ACGGACCAGC	CGCTCCCAGA	CGCCGACCAC	5100
ACCTGGTACA	CGGATGGAAG	CAGTCTCTTA	CAAGAGGGAC	AGCGTAAGGC	GGGAGCTGCG	5160
GTGACCACCG	AGACCGAGGT	AATCTGGGCT	AAAGCCCTGC	CAGCCGGGAC	ATCCGCTCAG	5220
CGGGCTGAAT	TATTAGCACT	CACCCAGGCC	CTAAGGATGG	CAGAAGGTAA	GAAGCTAAAT	5280
GTTTATACTG	ATAGCCGTTA	TGCTTTTGCT	ACTGCCCCATA	TCCATGGAGA	AATATACAGA	5340
AGGCGTGGGT	TGCTCACATC	AGAAGGCAAA	GAGATCAAAA	ATAAAGACGA	GATCTTGGCC	5400
CTACTAAAAG	CCCTCTTTCT	GCCCCAAAAG	CTTAGCATAA	TCCATTGTCC	AGGACATCAA	5460
AAGGGACACA	GCGCCGAGGC	TAGAGGCAAC	GAGCTGGCTG	ACCAAGCGGC	CCGAAAGGCA	5520
GCCATCACAG	AGACTCCAGA	CACCTCTACC	CTCCTCATAG	AAAATTATC	ACCCCTACAC	5580
TCAGAACATT	TTCATTACAC	AGTACTGAT	ATAAAGGACC	TAACCAAGTT	GGGGGCCATT	5640
TATGATAAAA	CAAAGAAGTA	TTGGGTCTAC	CAAGGAAAAC	CTGTGATGCC	TGACCACTTT	5700
ACTTTTGAAT	TATTAGACTT	TCTTCATCAG	CTGACTCACC	TCAGCTTCTC	AAAAATGAAG	5760
GCTCTCCTAG	AGAGAAGCCA	CAGTCCCTAC	TACATGCTGA	ACCGGGATCG	AACACTCAAA	5820
AATATCACTG	AGACCTGCAA	AGCTTGTGCA	CAAGTCAACG	CCAGCAAGTC	TGCCGTTAAA	5880
CAGGGAACCTA	GGGTCCGCGG	GCATCGGCCC	GGCAGCTATT	GGGAGATCGA	TTTCACCGAG	5940
ATAAAGCCCG	GATTGTATGG	CTATAAATAT	CTTCTAGTTT	TTATAGATAC	CTTTTCTGGC	6000
TGGATAGAAG	CCTTCCCAAC	CAAGAAAGAA	ACCGCAAGG	TCTGTAACCA	GAAGCTACTA	6060
GAGGAGATCT	TCCCCAGGTT	CGGCATGCCT	CAGGTATTGG	GAAGTGACAA	TGGGCCTGCC	6120
TTCGTCTCCA	AGGTGAGTCA	GACAGTGGCC	GATCTGTTGG	GGATTGATTG	GAAATTACAT	6180
TGTGCATACA	GACCCCAAAG	CTCAGGCCAG	TAGAAAAGAA	TGAATAGAAC	CATCAAGGAG	6240
ACTTTAACTA	AATTAACGCT	TGCAACTGGC	TCTAGAGACT	GGGTGCTCCT	ACTCCCCTTA	6300
GCCCTGTACC	GAGCCCCGAA	CACGCCGGGC	CCCCATGGCC	TCACCCCAT	TGAGATCTTA	6360
TATGGGGCAC	CCCCGCCCTT	TGTAACTTTC	CCTGACCCTG	ACATGACAAG	AGTTACTAAC	6420
AGCCCCCTCT	TCCAAGCTCA	CTTACAGGCT	CTCTACTTAG	TCCAGCACGA	AGTCTGGAGA	6480
CCTCTGGCGG	CAGCCTACCA	AGAACAACCTG	GACCGACCGG	TGGTACCTCA	CCCTTACCGA	6540
GTCCGGCACA	CAGTGTGGGT	CCGCCGACAC	CAGACTAAGA	ACCTAGAACC	TCGCTGGAAA	6600
GGACCTTACA	CAGTCTTGCT	GACCACCCCC	ACCGCCCTCA	AAGTAGACGG	CATCGCAGCT	6660
TGGATACACG	CCGCCACGCT	GAAGGCTGCC	GACCCCGGGG	GTGGACCATC	CTCTAGACTG	6720
ACATGGCGCG	TTCAACGCTC	TCAAAACCCC	TTAAAAATAA	GGTTAAACCCG	CGAGGCCCCC	6780
TAATCCCCTT	AATTCTTCTG	ATGCTCAGAG	GGGTCAAGTAC	TGCTTCGCCC	GGCTCCAGTG	6840
CGGCCAGGCC	GGCCACCATG	AAAACATTTA	ACATTTCTCA	ACAAGATCTA	GAATTAGTAG	6900
AAGTAGCGAC	AGAGAAGATT	ACAATGCTTT	ATGAGATAAA	TAAACATCAT	GTGGGAGCGG	6960
CAATTCGTAC	GAAAACAGGA	GAAATCATTT	CGGCAGTACA	TATTGAAGCG	TATATAGGAC	7020
GAGTAACTGT	TTGTGCAGAA	GCCATTGCGA	TTGTAGTGC	AGTTTCAAT	GGACAAAAGG	7080
ATTTTGACAC	GATTGTAGCT	GTTAGACACC	CTTATTCTGA	CGAAGTAGAT	AGAAGTATTC	7140
GAGTGTTAAG	TCCTTGTTGGT	ATGTGTAGGG	AGTTGATTTT	AGACTATGCA	CCGATTGTTT	7200
TTGTGTTAAT	AGAAATGAAT	GGCAAGTTAG	TCAAACTAC	GATTGAAGAA	CTCATTCCAC	7260
TCAATATATC	CCGAAATTAA	AAGTTTATAC	ACCAAGCTTA	TCAATTC		7308

Figure 8. hCMV+intronkaSD Sequence

1

AGATCTCCCG	ATCCCTATG	GTCGACTCTC	AGTACAATCT	GCTCTGATGC	CGCATAGTTA	60
AGCCAGTATC	TGCTCCCTGC	TTGTGTGTTG	GAGGTCGCTG	AGTAGTCCGC	GAGCAAAATT	120
TAAGCTACAA	CAAGGCAAGG	CTTGACCGAC	AATTGCATGA	AGAATCTGCT	TAGGGTTAGG	180
CGTTTTGCGC	TGCTTCGCGA	TGTACGGGCC	AGATATACGC	GTGACATTG	ATTATTGACT	240
AGTTATTAAT	AGTAATCAAT	TACGGGGTCA	TTAGTTCATA	GCCCATATAT	GGAGTTCGCG	300
GTTACATAAC	TTACGGTAAA	TGCCCCGCCT	GGGTGACCGC	CCAACGACCC	CCGCCCATTG	360
ACGTCAATAA	TGACGTATGT	TCCCATAGTA	ACGCCAATAG	GGACTTTCCA	TTGACGTCAA	420
TGGGTGGACT	ATTTACGGTA	AACTGCCCCAC	TTGGCAGTAC	ATCAAGTGTA	TCATATGCCA	480
AGTACGCCCC	CATTGACGT	CAATGACGGT	AAATGGCCCC	CCTGGCATT	TGCCCAGTAC	540
ATGACCTTAT	GTTAGCTTTC	TACTTGGCAG	TACATCTACG	TATTAGTCAT	CGCTATTACC	600
ATGGTGATGC	GGTTTTGGCA	GTACATCAAT	GGGCGTGGAT	AGCGGTTTGA	CTCACGGGGA	660
TTTCCAAGTC	TCCACCCCAT	TGACGTCAAT	GGGAGTTTGT	TTTGGCACCA	AAATCAACGG	720
GACTTTCCAA	AATGTCGTAA	CAACTCCGCC	CCATTGACGC	AAATGGGCGG	TAGGCGTGTA	780
CGGTGGGAGG	TCTATATAAG	CAGAGCTCTC	TGGCTAACCT	GAGAACCCAC	TGCTTAACTG	840
GCTTATCGAA	ATGTCGACTG	AGAACTTCAG	GGTGAGTTTG	GGGACCCTTG	ATTGTTCTTT	900
CTTTTTCGCT	ATTGTAAAAT	TCATGTTATA	TGGAGGGGGC	AAAGTTTTCA	GGGTGTTGTT	960
TAGAATGGGA	AGATGTCCCT	TGTATCACCA	TGGACCCTCA	TGATAATTTT	GTTTCTTTTCA	1020
CTTTCTACTC	TGTTGACAAC	CATTGTCTCC	TCTTATTTTC	TTTTCATTTT	CTGTAACCTT	1080
TTCGTTAAAC	TTTAGCTTGC	ATTTGTAACG	AATTTTTTAA	TTCACTTTTG	TTTATTGTGC	1140
AGATTGTAAG	TACTTTCTCT	AATCACTTTT	TTTTCAAGGC	AATCAGGGTA	TATTATATTG	1200
TACTTCAGCA	CAGTTTTAGA	GAACAATTGT	TATAATTAAA	TGATAAGGTA	GAATATTTCT	1260
GCATATAAAT	TCTGGCTGGC	GTGGAAATAT	TCTTATTGGT	AGAAACAAC	ACATCCTGGT	1320
CATCATCCTG	CCTTTCTCTT	TATGGTTACA	ATGATATACA	CTGTTTGAGA	TGAGGATAAA	1380
ATACTCTGAG	TCCAAACCGG	GCCCCCTCTG	TAACCATGTT	CATGCCTTCT	TCTTTTTTCT	1440
ACAGCTCCTG	GGCAACGTGC	TGGTTGTTGT	GCTGTCTCAT	CATTTTGGCA	AGAATTGGCC	1500
GCAAGCTTCT	GCCAGATCGT	TCTGTGTTGT	CTCTGTCTGA	CTGTGTTTCT	GTATTTGTCT	1560
GAGAATATGG	GCCAGACTGT	TACCATCCCC	TTAAGTTTGA	CCTTAGGTCA	CTGGAATAAG	1620
GTCGAGCGGA	TGCCTCACAA	CCAGTCGGTA	GATGTCAAGA	AGAGACGTTG	GGTTACCTTC	1680
TGCTCTGCAG	AATGGCCAAC	CTTTAACGTC	GGATGGCCGC	GAGACGGCAC	CTTTAACCGA	1740
GACCTCAAGT	CCCAGGTTAA	GATCAAGGTC	TTTTACCTG	GCCCCGATGG	ACACCCAGAC	1800
CAGGTCCCCCT	ACATCGTGAC	CTGGGAAGCC	TTGGCTTTTG	ACCCCCCTCC	CTGGGTCAAG	1860
CCCTTTGTAC	ACCCTAAGCC	TCCGCCTCCT	CTTCCTCCAT	CCGCCCCGTC	TCTCCCCCTT	1920
GAACCTCCTC	GTTTCGACCCC	GCCTCGATCC	TCCCTTTTATC	CAGCCCTCAC	TCCTTCTCTA	1980
GGCGCCAAAC	CTAAACCTCA	AGTTCCTTCT	GACAGTGGGG	GGCCGCTCAT	CGACCTACTT	2040
ACAGAAGACC	CCCCGCCTTA	TAGGGACCCA	AGACACCCCT	CTTCCGACAG	GGACGGAAAT	2100
GGTGGAAGA	CGACCCCTGC	GGGAGAGGCA	CCGGACCCCT	CCCCAATGGC	ATCTCGCCTA	2160
CGTGGGAGAC	GGGAGCCCCC	TGTGGCCGAC	TCCACTACCT	CGCAGGCATT	CCCCCTCCGC	2220
GCAGGAGGAA	ACGGACAGCT	TCAATATCTG	CCGTCTCCTT	CTTCTGACCT	TTACAACCTG	2280
AAAAATAATA	ACCCTTCTTT	TTCTGAAGAT	CCAGGTAAAC	TGACAGCTCT	GATCGAGTCT	2340
GTTCTCATCA	CCCATCAGCC	EACCTGGGAC	GACTGTGAGC	AGCTGTTGGG	GACTCTGCTG	2400
ACCGGAGAAG	AAAAACAACG	GGTGCTCTTA	GAGGCTAGAA	AGGCGGTGCG	GGGCGATGAT	2460
GGGCGAGGTC	CTCAACTGCC	CAATGAAGTC	GATGCCGCTT	TTCCCTCGA	GCGCCGACAG	2520
TGGGATTACA	CCACCCAGGC	AGGACGAAC	CACCTAGTCC	ACTATCGCCA	GTTGCTCCTA	2580
GCGGGTCTCC	AAAACGCGGG	CAGAAGCCCC	ACCAATTTGG	CCAAGGTAAA	AGGAATAACA	2640
CAAGGGCCCA	ATGAGTCTCC	CTCGGCCTTC	CTAGAGAGAC	TTAAGGAAGC	CTATCGCAGG	2700
TACACTCCTT	ATGACCTTGA	GGACCCAGGG	CAAGAACTA	ATGTGTCTAT	GTCTTTTCA	2760
TGGCAGTCTG	CCCCAGACAT	TGGGAGAAAG	TTAGAGAGGT	TAGAAGATTT	AAAAACAAG	2820
ACGCTTGGAG	ATTTGGTTAG	AGAGGCAGAA	AAGATCTTTA	ATAAACGAGA	AACCCCGGAA	2880
GAAAGAGAGG	AACGTATCAG	GAGAGAAACA	GAGGAAAAAG	AAGAACGCCG	TAGGACAGAG	2940
GATGAGCAGA	AAGAGAAAGA	AAGAGATCGT	AGGAGACATA	GAGAGATGAG	CAAGCTATTG	3000
GCCACTGTCT	TTAGTGGACA	GAAACAGGAT	AGACAGGGAG	GAGAACGAAG	GAGGTCCCAA	3060
CTCGATCGCG	ACCAGTGTGC	CTACTGCAAA	GAAAAGGGGC	ACTGGGCTAA	AGATTGTCCC	3120
AAGAAACCAC	GAGGACCTCG	GGGACCAAGA	CCCCAGACCT	CCCTCCTGAC	CCTAGATGAC	3180
TAGGGAGGTC	AGGGTCAGGA	GCCCCCCCCT	GAACCCAGGA	TAACCTCAA	AGTCGGGGGG	3240
CAACCCGTCA	CCTTCTCGGT	AGATACTGGG	GCCCCAACCT	CCGTGCTGAC	CCAAAATCCT	3300
GGACCCCTAA	GTGATAAGTC	TGCCTGGGTC	CAAGGGGCTA	CTGGAGGAAA	GCGGTATCGC	3360
TGGACCACGG	ATCGCAAAGT	ACATCTAGCT	ACCGGTAAGG	TCACCCACTC	TTTCTCCCAT	3420
GTACCGAGCT	GTCCCTATCC	TCTGTTAGGA	AGAGATTTGC	TGACTAACT	AAAAGCCCAA	3480
ATCCACTTTG	AGGGATCAGG	AGCTCAGGTT	ATGGGACCAA	TGGGGCAGCC	CCTGCAAGTG	3540
TTGACCCTAA	ATATAGAAGA	TGAGCATCGG	CTACATGAGA	CCTCAAAGA	GCCAGATGTT	3600
TCTCTAGGGT	CCACATGGCT	GTCTGATTTT	CCTCAGGCCT	GGGCGGAAAC	CGGGGGCATG	3660
GGACTGGCAG	TTCCGCAAGC	TCCTCTGATC	ATACCTCTGA	AAGCAACCTC	TACCCCGCTG	3720
TCCATAAAAC	AATACCCCAT	GTCAACAAGAA	GCCAGACTGG	GGATCAAGCC	CCACATACAG	3780
AGACTGTTGG	ACCAGGGAAT	ACTGGTACCC	TGCCAGTCCC	CCTGGAACAC	GCCCCCTGCTA	3840
CCCGTTAAGA	AACCAGGGAC	TAATGATTAT	AGGCCTGTCC	AGGATCTGAG	AGAAGTCAAC	3900
AAGCGGGTGG	AAGACATCCA	CCCCACCGTG	CCCCAACCTT	ACAACCTCTT	GAGCGGGCTC	3960
CCACCGTCCC	ACCAGTGGTA	CACGTGTGCT	GATTTAAAGG	ATGCCCTTTT	CTGCCCTGAGA	4020
CTCCACCCCA	CCAGTCAGCC	TCTCTTCGCC	TTTGAGTGGA	GAGATCCAGA	GATGGGAATC	4080

Figure 8. hCMV+intronkaSD Sequence

2

TCAGGACAAT	TGACCTGGAC	CAGACTCCCA	CAGGGTTTCA	AAAACAGTCC	CACCCGTGTTT	4140
GATGAGGCAC	TGCACAGAGA	CCTAGCAGAC	TTCCGGATCC	AGCACCAGAG	CTTGATCCTG	4200
CTACAGTACG	TGGATGACTT	ACTGCTGGCC	GCCACTTCTG	AGCTAGACTG	CCAAACAAGGT	4260
ACTCGGGCCC	TGTTACAAAC	CCTAGGGAAC	CTCGGGTATC	GGGCCTCGGC	CAAGAAAGCC	4320
CAAAATTGCC	AGAAAAGGT	CAAGTATCTG	GGGTATCTTC	TAAAAGAGGG	TCAGAGATGG	4380
CTGACTGAGG	CCAGAAAAGA	GACTGTGATG	GGGCAGCCTA	CTCCGAAGAC	CCCTCGACAA	4440
CTAAGGGAGT	TCCTAGGGAC	GGCAGGCTTC	TGTCGCCTCT	GGATCCCTGG	GTTTGCAGAA	4500
ATGGCAGCCC	CCTTGTAACC	TCTCACCAAA	ACGGGGACTC	TGTTTAATTG	GGGCCCAGAC	4560
CAACAAGAGG	CCTATCAAGA	AATCAAGCAA	GCTCTTCTAA	CTGCCCCAGC	CCTGGGGTTG	4620
CCAGATTTGA	CTAAGCCCTT	TGAACCTTTT	GTCGACGAGA	AGCAGGGCTA	CGCCAAAGGT	4680
GTCTTAACGC	AAAAACTGGG	ACCTTGGCGT	CGGCCGGTGG	CCTACCTGTC	CAAAAAGCTA	4740
GACCCAGTAG	CAGCTGGGTG	GCCCCCTTGC	CTACGGATGG	TAGCAGCCAT	TGCCGTACTG	4800
ACAAAGGATG	CAGCAAGCT	AACCATGGGA	CAGCCACTAG	TCATTCTGGC	CCCCCATGCA	4860
GTAGAGGCAC	TAGTCAAAAC	ACCCCCGAC	CGCTGGCTTT	CCAACGCCCG	GATGACTCAC	4920
TATCAGGCCT	TGCTTTTGGA	CACGGACCGG	GTCCAGTTCG	GACCGGTGGT	AGCCCTGAAC	4980
CCGGCTACGC	TGCTCCCACT	GCCTGAGGAA	GGGCTGCAAC	ACAACCTGCCT	TGATATCCTG	5040
GCCGAAGCCC	ACGGAACCCG	ACCCGACCTA	ACGGACCAGC	CGCTCCCAGA	CGCCGACCAC	5100
ACCTGGTACA	CGGATGGAAG	CAGTCTCTTA	CAGGTGGGAC	AGCGTAAGGC	GGGAGCTGCG	5160
GTGACCACCG	AGACCGAGGT	AATCTGGGCT	AAAGCCCTGC	CAGCCGGGAC	ATCCGCTCAG	5220
CGGGCTGAAC	TGATAGCACT	CACCCAGGCC	CTAAAGATGG	CAGAAGGTAA	GAAGCTAAAT	5280
GTTTATACTG	ATAGCCGTTA	TGCTTTTGCT	ACTGCCCAT	TCCATGGAGA	AATATACAGA	5340
AGGCGTGGGT	TGCTCACATC	AGAAGGCAAA	CAGATCAAAA	ATAAAGACGA	GATCTTGGCC	5400
CTACTAAAAG	CCCTCTTTCT	GCCCAAAAGA	CTTAGCATAA	TCCATTGTCC	AGGACATCAA	5460
AAGGGACACA	GCCCGAGGGC	TAGAGGCAAC	CGGATGGCTG	ACCAAGCGGC	CCGAAAGGCA	5520
GCCATCACAG	AGACTCCAGA	CACCTCTACC	CTCCTCATAG	AAAATTCATC	ACCCTACACC	5580
TCAGAACATT	TTCATTACAC	AGTGACTGAT	ATAAAGGACC	TAACCAAGTT	GGGGGCCATT	5640
TATGATAAAA	CAAAGAAGTA	TTGGGTCTAC	CAAGGAAAAC	CTGTGATGCC	TGACCAGTTT	5700
ACTTTTGAAT	TATTAGACTT	TCCTCATCAG	CTGACTCACC	TCAGCTTCTC	AAAAATGAAG	5760
GCTCTCCTAG	AGAGAAGCCA	CAGTCCCTAC	TACATGCTGA	ACCGGGATCG	AACACTCAAA	5820
AATATCACTG	AGACCTGCAA	AGCTTGTGCA	CAGATCAAAC	CCAGCAAGTC	TGCCGTAAAA	5880
CAGGGAACCTA	GGGTCCGCGG	GCATCGGCCC	GGCACTCATT	GGGAGATCGA	TTTACCCGAG	5940
ATAAAGCCCCG	GATTGTATGG	CTATAAATAT	CTTCTAGTTT	TTATAGATAC	CTTTTCTGGC	6000
TGGATAGAAG	CCTTCCCAAC	CAAGAAAGAA	ACCGCCAAGG	TCGTAACCAA	GAAGCTACTA	6060
GAGGACATCT	TCCCCAGGTT	CGGCATGCCT	CAGGTATTGG	GAAGTACAA	TGGGCCTGCC	6120
TTCGTCTCCA	AGGTGAGTCA	GACAGTGGCC	GATCTGTTGG	GGATTGATTG	GAAATTACAT	6180
TGTGCATACA	GACCCCCAAG	CTCAGGCCAG	GTAGAAAGAA	TGAATAGAAC	CATCAAGGAG	6240
ACTTTAACTA	AATTAACGCT	TGCAACTGGC	TCTAGAGACT	GGGTGCTCCT	ACTCCCCCTA	6300
GCCCTGTACC	GAGCCCGCAA	CACGCCGGGC	CCCCATGGCC	TCACCCCAT	TGAGATCTTA	6360
TATGGGGCAC	CCCCGCCCTT	TGTAAACTTC	CCTGACCCTG	ACATGACAAG	AGTTACTAAC	6420
AGCCCCCTCTC	TCCAAGCTCA	ETTACAGGCT	CTCTACTTAG	TCCAGCACGA	AGTCTGGAGA	6480
CCTCTGGCGG	CAGCCTACCA	AGAACAACCTG	GACCGACCGG	TGGTACCTCA	CCCTTACCGA	6540
GTCCGGCACA	CAGTGTGGGT	CCGCCGACAC	CAGACTAAGA	ACCTAGAACC	TCGCTGGAAA	6600
GGACCTTACA	CAGTCTTGCT	GACCACCCCC	ACCGCCCTCA	AAGTAGACGG	CATCGCAGCT	6660
TGGATACACG	CCGCCCCACGT	GAAGGCTGCC	GACCCCGGGG	GTGGACCATC	CTCTAGACTG	6720
ACATGGCGCG	TTCAACGCTC	TCAAAACCCC	TTAAAAATAA	GGTTAACCCG	CGAGGCCCCC	6780
TAATCCCCCTT	AATTCTTCTG	ATGCTCAGAG	GGGTCAGTAC	TGCTTCGCCC	GGCTCCAGTG	6840
CGGCCAGCC	GGCCACCATG	AAAACATTTA	ACATTTCTCA	ACAAGATCTA	GAATTAGTAG	6900
AAGTAGCGAC	AGAGAAGATT	ACAATGCTTT	ATGAGGATAA	TAAACATCAT	GTGGGAGCGG	6960
CAATTCGTAC	GAAAACAGGA	GAAATCATTT	CGGCAGTACA	TATTGAAGCG	TATATAGGAC	7020
GAGTAACTGT	TTGTGCAGAA	GCCATTGCGA	TTGGTAGTGC	AGTTTCGAAT	GGACAAAAGG	7080
ATTTTGACAC	GATTGTAGCT	GTTAGACACC	CTTATTCTGA	CGAAGTAGAT	AGAAGTATTC	7140
GAGTGGTAAG	TCCTTGTGGT	ATGTGTAGGG	AGTTGATTTC	AGACTATGCA	CCAGATTGTT	7200
TTGTGTTAAT	AGAAATGAAT	GGCAAGTTAG	TCAAACTAC	GATTGAAGAA	CTCATTCAC	7260
TCAAATATAC	CCGAAATTAA	AAGTTTTTACC	ACCAAGCTTA	TCGAATTC		7308



Figure 9. FBdelPASAF Sequence

CATATGCGGT	GTGAAATACC	GCACAGATGC	GTAAGGAGAA	AATACCGCAT	CAGGCGCCAT	60
TCGCCATTCA	GGCTGCGCAA	CTGTTGGGAA	GGGCGATCGG	TGCGGGCCTC	TTGCTATTA	120
CGCCAGCTGG	CGAAAGGGGG	ATGTGCTGCA	AGGCGATTAA	GTTGGGTAAAC	GCCAGGGTTT	180
TCCCAGTCAC	GACGTTGTAA	AACGACGGCC	AGTGAATTCC	GATTAGTTCA	ATTTGTTAAA	240
GACAGGATCT	CAGTAGTCCA	GGCTTTAGTC	CTGACTCAAC	AATACCACCA	GCTAAAACCA	300
CTAGAATACG	AGCCACAATA	AATAAAAGAT	TTTATTTAGT	TTCCAGAAAA	AGGGGGGAAT	360
GAAAGACCCC	ACCAAATTGC	TTAGCCTGAT	AGCCGCAGTA	ACGCCATTTT	GCAAGGCATG	420
GAAAAATACC	AAACCAAGAA	TAGAGAAGTT	CAGATCAAGG	GCGGGTACAC	GAAAACAGCT	480
AACGTTGGTG	CAACAGGAT	ATCTGCGGTG	AGCAGTTTTC	GCCCCGGCCC	GGGGCCAAGA	540
ACAGATGGTC	ACCGCGGTTT	GGCCCCGGCC	CGGGGCCAAG	AACAGATGGT	CCCCAGATAT	600
GGCCCAACCC	TCAGCAGTTT	CTTAAGACCC	ATCAGATGTT	TCCAGGCTCC	CCCAAGGACC	660
TGAAATGACC	CTGTGCCTTA	TTGAATTAA	CCAATCAGCC	TGCTTCTCGC	TTCTGTTCCG	720
CGCCTTCTGC	TTCCCGAGCT	CTATAAAGA	GCTCACAACC	CCTCACTCGG	CGCGCCAGTC	780
CTCCGATAGA	CTGGATCGCC	CGGGTACCCG	TGTATCCAAT	AAATCCTCTT	GCTGTTGCAT	840
CCGACTCGTG	GTCTCGCTGT	TCCTTGGGAG	GGTCTCCTCA	GAGTGATTGA	CTACCCGTCT	900
CGGGGGTCTT	TCATTTGGGG	GCTCGTCCGG	GATCTGGAGA	CCCCTGCCCA	GGGACCACCG	960
ACCCACACC	GGGAGGTAAG	CTGGCCAAGA	TCTTATATGG	GGCACCCCCG	CCCCTTGTA	1020
ACTTCCCTGA	CCCTGACATG	ACCAGAGTTA	CTAACAGCCC	CTCTCTCCAA	GCTCACTTAC	1080
AGGCTCTCTA	CTTAGTCCAG	CACGAAGTTT	GGAGACCACT	GGCGGCAGCT	TACCAAGAAC	1140
AACTGGACCG	GCCGGTGGTG	CCTCACCCCT	ACCGGGTCGG	CGACACAGTG	TGGGTCCGCC	1200
GACATCAAA	CAAGAACCTA	GAACCTCGCT	GGAAAGGACC	TTACACAGTC	CTGCTGACCA	1260
CCCCACCGC	CCTCAAGTA	GACGGTATCG	CAGCTTGGAT	ACACGCAGCC	CACGTAAAGG	1320
CGGCCGACAC	CGAGAGTGGG	CCATCCTCTG	GACGGACATG	GCGCGTTCAA	CGCTCTCAA	1380
ACCCCTCTAA	GATAAGATTA	ACCCGTGGAA	GCCCTTAATA	GTCATGGGAG	TCCTGTTAGG	1440
AGTAGGGATG	GCAGAGAGCC	CCCATCAGGT	CTTTAATGTA	ACCTGGAGAG	TCACCAACCT	1500
GATGATCTGG	CGTAGCCGCA	ATGCCACCTC	CTCCTGGGA	ACTGTACAAG	ATGCCTTCCC	1560
AAAATTATAT	TTTGATCTAT	GTGATCTGGT	CGGAGGAGAG	TGGGACCCTT	CAGACCAGGA	1620
ACCGTATGTC	GGGTATGGCT	GCAAGTACCC	CGCAGGGAGA	CAGCGGACCC	GGACTTTTGA	1680
CTTTTACGTG	TGCCCTGGGC	ATACCGTAAA	GTCGGGGTGT	GGGGGACCAG	GAGAGGGCTA	1740
CTGTGGTAAA	TGGGGGTGTG	AAACCACCGG	ACAGGCTTAC	TGGAAGCCCA	CATCATCGTG	1800
GGACCTAATC	TCCCTTAAGC	GCGGTAACAC	CCCCTGGAC	ACGGGATGCT	CTAAAGTTGC	1860
CTGTGGCCCC	TGCTACGACC	TCTCCAAAGT	ATCCAATTCC	TTCCAAGGGG	CTACTCGAGG	1920
GGGCAGATGC	AACCTCTTAG	TCCTAGAATT	CACTGATGCA	GGAAAAAAGG	CTAAGTGGGA	1980
CGGGCCCAA	TCGTGGGGAC	TGAGACTGTA	CCGGACAGGA	ACAGATCCTA	TTACCATGTT	2040
CTCCCTGACC	CGGCAGGTCC	TTAATGTGGG	ACCTGAGATC	CCCATAGGGC	CCAACCCAGT	2100
ATTACCCGAC	CAAAGACTCC	CTTCTCACC	AATAGAGATT	GTACCGGCTC	CACAGCCACC	2160
TAGCCCCCTC	AATACCAGTT	ACCCCTTTC	CACTACCAGT	ACACCTCAA	CCTCCCTAC	2220
AAGTCCAAGT	GTCCACAGC	CACCCCAGG	AACCTGGAGT	AGACTACTAG	CTCTAGTCAA	2280
AGGAGCCTAT	CAGGCGCTTA	ACCTACCAA	TCCCGACAAG	ACCCAAGAAT	GTTGGCTGTG	2340
CTTAGTGTG	GGACCTCCTT	ATTACGAAGG	AGTAGCGGTC	GTGGGCACTT	ATACCAATCA	2400
TTCCACCGCT	CCGGCCAACT	GTACGGCCAC	TTCCCAACAT	AAGCTTACCC	TATCTGAAGT	2460
GACAGGACAG	GGCCTATGCA	TGGGGGCAGT	ACCTAAACT	CACGAGGCTT	TATGTAACAC	2520
CACCCAAAGC	GCCGGCTCAG	GATCTACTA	CCTTGACGA	CCCGCCGGAA	CAATGTGGGC	2580
TTGCAGCACT	GGATTGACTC	CCTGCTTGTC	CACCACGGTG	CTCAATCTAA	CCACAGATTA	2640
TTGTGTATTA	GTTGAACCTT	GGCCAGAGT	AATTTACCAC	TCCCCCGATT	ATATGTATGG	2700
TCAGCTTGAA	CAGCGTACCA	AATATAAAG	AGAGCCAGTA	TCATTGACCC	TGGCCCTTCT	2760
ACTAGGAGGA	TTAACCATGG	GAGGGATTGC	AGCTGGAAATA	GGGACGGGGA	CCACTGCTCT	2820
AATTAAACC	CAGCAGTTTG	AGCAGCTTCA	TGCCGCTATC	CAGACAGACC	TCAACGAAGT	2880
CGAAAAGTCA	ATTACCAACC	TAGAAAAGTC	ACTGACCTCG	TTGTCTGAAG	TAGTCTTACA	2940
GAACCGCAGA	GGCCTAGATT	TGCTATTCTT	AAAGGAGGGA	GGTCTCTGCG	CAGCCCTAAA	3000
AGAAGAATGT	TGTTTTTATG	CAGACCACAC	GGGGCTAGTG	AGAGACAGCA	TGGCCAAATT	3060
AAGAGAAAGG	CTTAATCAGA	GACAAAAACT	ATTTGAGACA	GGCCAAGGAT	GGTTCGAAGG	3120
GCTGTTTAAT	AGATCCCCCT	GGTTTACCAC	CTTAATCTCC	ACCATCATGG	GACCTCTAAT	3180
AGTACTCTTA	CTGATCTTAC	TCTTTGGACC	TTGCATCTCT	AATCGATTAG	TTCAATTTGT	3240
TAAAGACAGG	ATCTCAGTAG	TCCAGGCTTT	AGTCTGACT	CAACAATACC	ACCAGCTAAA	3300
GCCTATAGAG	TACGAGCCAT	AGGGCGCCTA	GTGTTGACAA	TTAATCATCG	GCATAGTATA	3360
CGGCATAGTA	TAATACGACT	CACATATAGGA	GGGCCACCAT	GGCCAAGTTG	ACCAGTGCCG	3420
TTCCGGTGCT	CACCGCGCG	GACGTCGCGG	GAGCGGTGCA	GTCTTGACC	GACCGGCTCG	3480
GGTCTCCCCG	GGACTTCGTG	GAGGACGACT	TCCGCGGTGT	GGTCCGGGAC	GACGTGACCC	3540
TGTTTCATCAG	CGCGGTCCAG	GACCAGGTGG	TGCCGGACAA	CACCTTGCC	TGGGTGTGGG	3600
TGCGCGGCCCT	GGACGAGCTG	TACGCCGAGT	GGTCCGAGGT	CGTGTCCACG	AACCTCCGGG	3660
ACCGTCCCG	ACCGGCCATG	ACCGAGATCG	GCGAGCAGCC	GTGGGGGCGG	GAGTTCGCCC	3720
TGCGCGACCC	GGCCGGCAAC	TGCGTGCACT	TCCGTGCGGA	GGAGCAGGAC	TGANNNNCGG	3780
ACCGGTCGAC	TTGTTAACTT	GTTTATTGCA	GCTTATAATG	GTTACAAATA	AAGCAATAGC	3840
ATCACAAATT	TCACAAATAA	AGCATTTTTT	TCATCTGCATT	CTAGTTGTGG	TTGTGCCAAA	3900
CTCATCAATG	TATCTTATCA	TGTCTGGATC	CAGATCTGGG	CCCATGCGGC	CCCGGATCGA	3960
TNNNNAATG	TAGGCAAAAAG	GCCAGCAAAA	GCCAGGAAAC	CGTAAAAAGG	CCGCGTTGCT	4020
GGCGTTTTTC	CATAGGCTCC	GCCCCCTGA	CGAGCATCAC	AAAAATCGAC	GCTCAAGTCA	4080

Figure 9. FBdelPASAF Sequence

2

GAGGTGGCGA	AACCCGACAG	GACTATAAAG	ATACCAGGCG	TTTCCCCCTG	GAAGCTCCCT	4140
CGTGCGCTCT	CCTGTTCCGA	CCCTGCCGCT	TACCGGATAC	CTGTCCGCCT	TTCTCCCTTC	4200
GGGAAGCGTG	GCGCTTTCTC	AATGCTCAG	CTGTAGGTAT	CTCAGTTCGG	TGTAGGTCGT	4260
TCGCTCCAAG	CTGGGCTGTG	TGCACGAACC	CCCCGTTTCC	CCCGACCGCT	GCGCCTTATC	4320
CGGTAACAT	CGTCTTGAGT	CCAACCCGGT	AAGACACGAC	TTATCGCCAC	TGGCAGCAGC	4380
CACTGGTAAC	AGGATTAGCA	GAGCGAGGTA	TGTAGGCGGT	GCTACAGAGT	TCTTGAAGTG	4440
GTGGCCTAAC	TACGGCTACA	CTAGAAGGAC	AGTATTTGGT	ATCTGCGCTC	TGCTGAAGCC	4500
AGTTACCTTC	GGAAAAAGAG	TTGGTAGCTC	TTGATCCGGC	AAACAAACCA	CCGCTGGTAG	4560
CGGTGGTTT	TTTGTTTGCA	AGCAGCAGAT	TACGCGCAGA	AAAAAAGGAT	CTCAAGAAGA	4620
TCCTTTGATC	TTTTCTACGG	GGTCTGACGC	TCAGTGGAAC	GAAAACTCAC	GTTAAGGGAT	4680
TTTGGTCAAT	AGATTATCAA	AAAGGATCTT	CACCTAGATC	CTTTTAAATT	AAAAATGAAG	4740
TTTTAAATCA	ATCTAAAGTA	TATATGAGTA	AACTTGGTCT	GACAGTTACC	AATGCTTAAT	4800
CAGTGAGGCA	CCTATCTCAG	CGATCTGTCT	ATTTTCGTTCA	TCCATAGTTG	CCTGACTCCC	4860
CGTCGTGTAG	ATAACTACGA	TACGGGAGGG	CTTACCATCT	GGCCCCAGTG	CTGCAATGAT	4920
ACCGCGAGAC	CCACGCTCAC	CGGCTCCAGA	TTTATCAGCA	ATAAACCAGC	CAGCCGGAAG	4980
GGCCGAGCGC	AGAAGTGGTC	CTGCAACTTT	ATCCGCCTCC	ATCCAGTCTA	TTAATTGTTG	5040
CCGGGAAGCT	AGAGTAAGTA	GTTTCGCCAGT	TAATAGTTTG	CGCAACGTTG	TTGCCATTGC	5100
TACAGGCATC	GTGGTGTGAC	GCTCGTCGTT	TGGTATGGCT	TCATTCAGCT	CCGGTTCCCA	5160
ACGATCAAGG	CGAGTTACAT	GATCCCCCAT	GTTGTGCAAA	AAAGCGGTTA	GCTCCTTCGG	5220
TCCTCCGATC	GTTGTGAGAA	GTAAGTTGGC	CGCAGTGTTA	TCACTCATGG	TTATGGCAGC	5280
ACTGCATAAT	TCTCTTACTG	TCATGCCATC	CGTAAGATGC	TTTTCTGTGA	CTGGTGAGTA	5340
CTCAACCAAG	TCATTCTGAG	AATAGTGTAT	GCGGCGACCG	AGTTGCTCTT	GCCCCGCGTC	5400
AATACGGGAT	AATACCGCGC	CACATAGCAG	AACTTTAAAA	GTGCTCATCA	TTGGAAAACG	5460
TTCTTCGGGG	CGAAAACCTC	CAAGGATCTT	ACCGCTGTTG	AGATCCAGTT	CGATGTAACC	5520
CACTCGTGCA	CCCAACTGAT	CTTCAGCATC	TTTTACTTTC	ACCAGCGTTT	CTGGGTGAGC	5580
AAAAACAGGA	AGGCAAAATG	CCGCAAAAAA	GGGAATAAGG	GCGACACGGA	AATGTTGAAT	5640
ACTCATACTC	TTCTTTTTC	AATATTATTG	AAGCATTAT	CAGGGTTATT	GTCTCATGAG	5700
CGGATACATA	TTTGAATGTA	TTTAGAAAAA	TAAACAAATA	GGGGTTCCGC	GCACATTTC	5760
CCGAAAAGTG	CCACCTGACG	TCTAAGAAAC	CATTATTATC	ATGACATTAA	CCTATAAAAA	5820
TAGGCGTATC	ACGAGGCCCT	TTCGTCTCGC	GCGTTTCGGT	GATGACGGTG	AAAACCTCTG	5880
ACACATGCAG	CTCCCGGAGA	CGGTCACAGC	TTGTCTGTAA	GCGGATGCCG	GGAGCAGACA	5940
AGCCCGTCAG	GGCGCGTCAG	CGGGTGTGG	CGGGTGTCCG	GGCTGGCTTA	ACTATGCGGC	6000
ATCAGAGCAG	ATTGTACTGA	GAGTGCAC				6028

Figure 10. FBdelPMOSAF Sequence

1

CATATGCGGT	GTGAAATACC	GCACAGATGC	GTAAGGAGAA	AATACCGCAT	CAGGCGCCAT	60
TCGCCATTCA	GGCTGCGCAA	CTGTTGGGAA	GGGCGATCGG	TGCGGGCCTC	TTGCTATTA	120
CGCCAGCTGG	CGAAAGGGGG	ATGTGCTGCA	AGGCGATTAA	GTGCGGTAAC	GCCAGGGTTT	180
TCCCAGTCAC	GACGTTGTAA	AACGACGGCC	AGTGAATTCC	GATTAGTTCA	ATTTGTTAAA	240
GACAGGATCT	CAGTAGTCCA	GGCTTTAGTC	CTGACTCAAC	AATACCACCA	GCTAAACCA	300
CTAGAATACG	AGCCACAATA	AATAAAAGAT	TTTATTTAGT	TTCCAGAAAA	AGGGGGGAAT	360
GAAAGACCCC	ACCAAATTGC	TTAGCCTGAT	AGCCGCAGTA	ACGCCATTTT	GCAAGGCATG	420
GAAAAATACC	AAACCAAGAA	TAGAGAAGTT	CAGATCAAGG	GCGGGTACAC	GAAAAACAGCT	480
AACGTTGGGC	CAAACAGGAT	ATCTGCGGTG	AGCAGTTTCG	GCCCCGGCCC	GGGGCCAAGA	540
ACAGATGGTC	ACCGCGGTTT	GGCCCCGGCC	CGGGGCCAAG	AACAGATGGT	CCCCAGATAT	600
GGCCCAACCC	TCAGCAGTTT	CTTAAGACCC	ATCAGATGTT	TCCAGGCTCC	CCCAAGGACC	660
TGAAATGACC	CTGTGCCTTA	TTTGAATTAA	CCAATCAGCC	TGCTTCTCGC	TTCTGTTTCG	720
GCGCTTCTGC	TTCCCGAGCT	CTATAAAAGA	GCTCAACAAC	CCTCACTCGG	CGCGCCAGTC	780
CTCCGATAGA	CTGAGTCGCC	CGGGTACCCG	TGTATCCAAT	AAATCCTCTT	GCTGTTGCAT	840
CCGACTCGTG	GTCTCGCTGT	TCCTTGGGAG	GGTCTCCTCA	GAGTGATTGA	CTACCCGTCT	900
CGGGGCTCTT	TCATTTGGGG	GCTCGTCCGG	GATCTGGAGA	CCCCTGCCCA	GGGACCACCG	960
ACCCACCCAC	CTGGGTAAG	CTGGCCAAGA	TCTTATATGG	GGCACCCTCG	CCCCTTGTA	1020
ACTTCCCTGA	CCCTGACATG	ACAAGAGTTA	CTAACAGCCC	CTCTCTCCAA	GCTCACTTAC	1080
AGGCTCTCTA	CTTAGTCCAG	CACGAAGTCT	GGAGACCTCT	GGCGGCAGCC	TACCAAGAAC	1140
AACTGGACCG	ACCGGTGGTA	CCTCACCCCT	ACCGAGTCGG	CGACACAGTG	TGGGTCCGCC	1200
GACACCAGAC	TAAGAACCTA	GAACCTCGCT	GGAAAGGACC	TTACACAGTC	CTGCTGACCA	1260
CCCCCACCGC	CCTCAAAGTA	GACGGCATCG	CAGCTTGGAT	ACACGCCGCC	CACGTGAAGG	1320
CTGCCGACCC	CGGGGGTGA	CCATCCTCTA	GACTGACATG	GCGCGTTCAA	CGCTCTCAA	1380
ACCCCTTAAA	AATAAGGTTA	ACCCGCGAGG	CCCCCTAATC	CCCTTAATTC	TTCTGATGCT	1440
CAGAGGGGTC	AGTACTGCTT	CGCCCGGCTC	CAGTCCTCAT	CAAGTCTATA	ATATCACCTG	1500
GGAGGTAACC	AATGGAGATC	GGGAGACGGT	ATGGGCAACT	TCTGGCAACC	ACCCCTCTGT	1560
GACCTGGTGG	CCTGACCTTA	CCCCAGATTT	ATGTATGTTA	GCCCACCATG	GACCATCTTA	1620
TTGGGGGCTA	GAATATCAAT	CCCCCTTTTC	TTCTCCCCCG	GGGCCCCCTT	GTTGCTCAGG	1680
GGGCAGCAGC	CCAGGCTGTT	CCAGAGACTG	CGAAGAACCT	TTAACCTCCC	TCACCCCTCG	1740
GTGCAACACT	GCCTGGAAAC	GACTCAAGCT	AGACCAGACA	ACTCAATAAT	CAATGAGGG	1800
ATTTTATGTT	TGCCCGGGGC	CCCACCGCCC	CCGAGAATCC	AAGTCATGTG	GGGGTCCAGA	1860
CTCCTTCTAC	TGTGCCTATT	GGGGCTGTGA	GACAACCGGT	AGAGCTTACT	GGAAGCCCTC	1920
CTCATCATGG	GATTTTCATCA	CAGTAAACAA	CAATCTCACC	TCTGACCAGG	CTGTCCAGGT	1980
ATGCAAAGAT	AATAAGTGGT	GCAACCCCTT	AGTTATTCGG	TTTACAGACG	CCGGGAGACG	2040
GGTTACTTCC	TGGACCACAG	GACATTACTG	GGGCTTACGT	TTGTATGTCT	CCGGACAAGA	2100
TCCAGGGCTT	ACATTTGGGA	TCCGACTCAG	ATACCAAAAT	CTAGGACCCC	GCGTCCCAAT	2160
AGGGCCAAAC	CCGTTCTCGG	CAGACCAACA	GCCACTCTCC	AAGCCCAAAC	CTGTTAAGTC	2220
GCCTTCAGTC	ACCAAACAC	CCAGTGGGAC	TCCTCTCTCC	CCTACCAAAC	TTCCACCGGC	2280
GGGAACGGAA	AATAGGCTGC	TAAACTTAGT	AGACGGAGCC	TACCAAGCCC	TCAACCTCAC	2340
CAGTCTTGAC	AAAACCCAAG	AGTGCTGGTT	GTGCTAGTA	GCGGGACCCC	CCTACTACGA	2400
AGGGTGTGCC	GTCTGGGTA	CTTACTCCAA	CCATACCTCT	GCTCCAGCCA	ACTGCTCCGT	2460
GGCCTCCCAA	CACAAGTTGA	CCCTGTCCGA	AGTGACCGGA	CAGGGACTCT	GCATAGGAGC	2520
AGTTCCCAAA	ACACATCAGG	CCCTATGTAA	TACCACCCAG	ACAAGCAGTC	GAGGGTCTTA	2580
TTATCTAGTT	GCCCTTACAG	GTACCATGTG	GGCTTGTAGT	ACCGGGCTTA	CTCCATGCAT	2640
CTCCACCACC	ATACTGAACC	TTACCACTGA	TTATTGTGTT	CTTGTCGAAC	TCTGGCCAAG	2700
AGTCACCTAT	CATTCCCCCA	GCTATGTTTA	CGGCTGTTT	GAGAGATCCA	ACCGACCAA	2760
AAGAGAACCG	GTGTCGTTAA	CCCTGGCCCT	ATTATTGGGT	GGACTAACCA	TGGGGGGAAT	2820
TGCCGCTGGA	ATAGGAACAG	GGACTACTGC	TCTAATGGCC	ACTCAGCAAT	TCCAGCAGCT	2880
CCAAGCCGCA	GTACAGGATG	ATCTCAGGGA	GGTTGAAAAA	TCAATCTCTA	ACCTAGAAAA	2940
GTCTCTCACT	TCCCTGTCTG	AAGTTGTCTT	ACAGAATCGA	AGGGGCCTAG	ACTTGTATT	3000
TCTAAAAGAA	GGAGGGCTGT	GTGCTGCTCT	AAAAGAAGAA	TGTTGCTTCT	ATGCGGACCA	3060
CACAGGACTA	GTGAGAGACA	GCATGGCCAA	ATTGAGAGAG	AGGCTTAATC	AGAGACAGAA	3120
ACTGTTTGAG	TCAACTCAAG	GATGGTTTGA	GGGACTGTTT	AACAGATCCC	CTTGGTTTAC	3180
CACCTTGATA	TCTACCATT	TGGGACCCCT	CATTGTACTC	CTAATGATTT	TGCTCTTCGG	3240
ACCTGCATT	CTTAATCGAT	TAGTTCAATT	TGTTAAAGAC	AGGATCTCAG	TAGTCCAGGC	3300
TTAGTCTCTG	ACTCAACAAT	ACCACCAGCT	AAAGCCTATA	GAGTACGAGC	CATAGGGCGC	3360
CTAGTGTTGA	CAATTAATCA	TCGGCATAGT	ATACGGCATA	GTATAATACG	ACTCACTATA	3420
GGAGGGCCAC	CATGGCCAAG	TTGACCAGTG	CCGTTCCGGT	GCTCACCCGC	CGCGACGTCG	3480
CCGGAGCGGT	CGAGTTCTGG	ACCGACCGGC	TCGGGTTCTC	CCGGGACTTC	GTGGAGGACG	3540
ACTTCGCCGG	TGTGGTCCGG	GACGACGTGA	CCCTGTTTCT	CAGCGCGGTC	CAGGACCAGG	3600
TGGTGCCGGA	CAACACCCCTG	GCCTGGGTGT	GGGTGCGCGG	CCTGGACGAG	CTGTACGCCG	3660
AGTGGTCCGA	GGTCGTGTCC	ACGAACCTTC	GGGACCGCTC	CGGGCCGGCC	ATGACCGAGA	3720
TCGGCGAGCA	GCCGTGGGGG	CGGGAGTTCTG	CCCTGCGCGA	CCCGCCGGGC	AACTGCGTGC	3780
ACTTCGTGGC	CGAGGAGCAG	GACTGANNNN	CGGACCGGTC	GACTTGTTAA	CTTGTGTTATT	3840
CGAGCTTATA	ATGGTTACAA	ATAAAGCAAT	AGCATCACAA	ATTTACACAA	TAAAGCATTT	3900
TTTTCATGTC	ATGTTAGTTG	TGGTTTGTCC	AAACTCATCA	ATGTATCTTA	TCATGTCTGG	3960
ATCCAGATCT	GGGCCCATGC	GGCCGCGGAT	CGATNNNNAC	ATGTGAGCAA	AAGGCCAGCA	4020
AAAGGCCAGG	AACCGTAAAA	AGCCCGCGTT	GCTGGCGTTT	TTCCATAGGC	TCCGCCCCC	4080

Figure 10. FBdelPMOSAF Sequence

TGACGAGCAT	CACAAAAATC	GACGCTCAAG	TCAGAGGTGG	CGAAACCCGA	CAGGACTATA	4140
AAGATAACCAG	GCGTTTCCCC	CTGGAAGCTC	CCTCGTGCGC	TCTCCTGTTC	CGACCCTGCC	4200
GCTTACCGGA	TACCTGTCCG	CCTTTCTCCC	TTCCGGAAGC	GTGGCGCTTT	CTCAATGCTC	4260
ACGCTGTAGG	TATCTCAGTT	CGGTGTAGGT	CGTTCGCTCC	AAGCTGGGCT	GTGTGCACGA	4320
ACCCCCGTT	CAGCCCGACC	GCTGCGCCTT	ATCCGGTAAC	TATCGTCTTG	AGTCCAACCC	4380
GGTAAGACAC	GACTTATCGC	CACTGGCAGC	AGCCACTGGT	AACAGGATTA	GCAGAGCGAG	4440
GTATGTAGGC	GGTGCTACAG	AGTTCTTGAA	GTGGTGGCCT	AACTACGGCT	ACACTAGAAG	4500
GACAGTATTT	GGTATCTGCG	CTCTGCTGAA	GCCAGTTACC	TTCCGAAAAA	GAGTTGGTAG	4560
CTCTTGATCC	GGCAAAACAA	CCACCGCTGG	TAGCCGTGGT	TTTTTTGTTC	GCAAGCAGCA	4620
GATTACGCGC	AGAAAAAAG	GATCTCAAGA	AGATCCTTTG	ATCTTTTCTA	CGGGGTCTGA	4680
CGCTCAGTGG	AACGAAAACT	CACGTTAAGG	GATTTTGGTC	ATGAGATTAT	CAAAAAGGAT	4740
CTTCACCTAG	ATCCTTTTAA	ATTAAAAATG	AAGTTTAAAA	TCAATCTAAA	GTATATATGA	4800
GTAAACTTGG	TCTGACAGTT	ACCAATGCTT	AATCAGTGAG	GCACCTATCT	CAGCGATCTG	4860
TCTATTTTCG	TCATCCATAG	TTGCCTGACT	CCCCGTCGTG	TAGATAACTA	CGATACGGGA	4920
GGGCTTACCA	TCTGGCCCCA	GTGCTGCAAT	GATACCGCGA	GACCCACGCT	CACCGGCTCC	4980
AGATTTATCA	GCAATAAACC	AGCCAGCCGG	AAGGGCCGAG	CGCAGAAGTG	GTCTTGCAAC	5040
TTTATCCGCC	TCCATCCAGT	CTATTAATTG	TTGCCGGGAA	GCTAGAGTAA	GTAGTTCGCC	5100
AGTTAATAGT	TTGCGCAACG	TTGTTGCCAT	TGCTACAGGC	ATCGTGGTGT	CACGCTCGTC	5160
GTTTGGTATG	GCTTCATTCA	GCTCCGGTTC	CCAACGATCA	AGGCGAGTTA	CATGATCCCC	5220
CATGTTGTGC	AAAAAAGCGG	TTAGCTCCTT	CGGTCTCCG	ATCGTTGTCA	GAAGTAAGTT	5280
GGCCGCAGTG	TTATCACTCA	TGGTTATGGC	AGCACTGCAT	AATTCTCTTA	CTGTCATGCC	5340
ATCCGTAAGA	TGCTTTTCTG	TGACTGGTGA	GTACTCAACC	AAGTCATTCT	GAGAATAGTG	5400
TATGCGGCGA	CCGAGTTGCT	CTTGCCCGGC	GTCAATACGG	GATAATACCG	CGCCACATAG	5460
CAGAACTTTA	AAAGTGCTCA	TCATTGGAAA	ACGTTCTTCG	GGGCGAAAAA	TCTCAAGGAT	5520
CTTACCGCTG	TTGAGATCCA	GTTTCGATGTA	ACCCACTCGT	GCACCCAAC	GATCTTCAGC	5580
ATCTTTTACT	TTCACCAGCG	TTTCTGGGTG	AGCAAAAAACA	GGAAGGCAAA	ATGCCGCAAA	5640
AAAGGGAATA	AGGGCGACAC	GGAAATGTTG	AATACTCATA	CTCTTCCTTT	TTCAATATTA	5700
TTGAAGCATT	TATCAGGGTT	ATTGTCTCAT	GAGCGGATAC	ATATTTGAAT	GTATTTAGAA	5760
AAATAAACAA	ATAGGGGTTC	CGCGCACATT	TCCCCGAAAA	GTGCCACCTG	ACGTCTAAGA	5820
AACCATTATT	ATCATGACAT	TAACCTATAA	AAATAGGCGT	ATCACGAGGC	CCTTTTCGTCT	5880
CGCGCGTTTC	GGTGATGACG	GTGAAAACCT	CTGACACATG	CAGCTCCCGG	AGACGGTCAC	5940
AGCTTGTCTG	TAAGCGGATG	CCGGGAGCAG	ACAAGCCCGT	CAGGGCGCGT	CAGCGGGTGT	6000
TGGCGGGTGT	CGGGGCTGGC	TTAACTATGC	GGCATCAGAG	CAGATTGTAC	TGAGAGTGCA	6060
C						6061

Figure 11. FBdelPGASAF Sequence

1

CATATGCGGT	GTGAAATACC	GCACAGATGC	GTAAGGAGAA	AATACCGCAT	CAGGCGCCAT	60
TCGCCATTCA	GGCTGCGCAA	CTGTTGGGAA	GGGCGATCGG	TGCGGGCCTC	TTCGCTATTA	120
CGCCAGCTGG	CGAAAGGGGG	ATGTGCTGCA	AGGCGATTAA	GTTGGGTAAAC	GCCAGGGTTT	180
TCCCAGTCAC	GACGTTGTAA	AACGACGGCC	AGTGAATTCC	GATTAGTTCA	ATTTGTTAAA	240
GACAGGATCT	CAGTAGTCCA	GGCTTTAGTC	CTGACTCAAC	AATACCACCA	GCTAAAACCA	300
CTAGAATACG	AGCCACAATA	AATAAAAGAT	TTTATTTAGT	TTCCAGAAAA	AGGGGGGAAT	360
GAAAGACCCC	ACCAAATTGC	TTAGCCTGAT	AGCCGCAGTA	ACGCCATTTT	GCAAGGCATG	420
GAAAAATACC	AAACCAAGAA	TAGAGAAGTT	CAGATCAAGG	GCGGGTACAC	GAAAAACAGCT	480
AACGTTGGGC	CAAACAGGAT	ATCTGCGGTG	AGCAGTTTTCG	GCCCCGGCCC	GGGGCCAAGA	540
ACAGATGGTC	ACCGCGGTTC	GGCCCCGGCC	CGGGGCCAAG	AACAGATGGT	CCCCAGATAT	600
GGCCCAACCC	TCAGCAGTTT	CTTAAGACCC	ATCAGATGTT	TCCAGGCTCC	CCCAAGGACC	660
TGAAATGACC	CTGTGCCTTA	TTTGAATTAA	CCAATCAGCC	TGCTTCTCGC	TTCTGTTTCG	720
GCGCTTCCGC	TTCCCGAGCT	CTATAAAAGA	GCTCAACACC	CCTCACTCGG	CGCGCCAGTC	780
CTCCGATAGA	CTGAGTCGCC	CGGGTACCCG	TGTATCCAAT	AAATCCTCTT	GCTGTTGCAT	840
CCGACTCGTG	GTCTCGCTGT	TCCTTGGGAG	GGTCTCCTCA	GAGTGATTGA	CTACCCGTCT	900
CGGGGGTCTT	TCATTTGGGG	GCTCGTCCGG	GATCTGGAGA	CCCCTGCCCA	GGGACCACCG	960
ACCCACCACC	GGGAGGTAAG	CTGGCCAAGA	TCCCTAAGGT	ACTCGGGTCA	GACAAATGGCC	1020
CGGCCTTTGT	TGCTCAGGTA	AGTCAGGGAC	TGGCCACTCA	ACTGGGGATA	AATTGGAAGT	1080
TACATTGTGC	GTATAGACCC	CAGAGCTCAG	GTCAGGTAGA	AAGAATGAAC	AGAACAATTA	1140
AAGAGACCTT	GACCAAATTA	GCCTTAGAGA	CCGGTGGAAA	AGACTGGGTG	ACCCTCCTTC	1200
CCTTAGCGCT	GCTTAGGGCC	AGGAATACCC	CTGGCCGGTT	TGGTTTAACT	CCTTATGAAA	1260
TTCTCTATGG	AGGACCACCC	CCCATACTTG	AGTCTGGAGA	AACTTTGGGT	CCCCATGATA	1320
GATTTCTCCC	TGTCTTATTT	ACTCACTTAA	AGGCTTTAGA	AATTGTAAGG	ACCCAAATCT	1380
GGGACCAGAT	CAAGAGGGTG	TATAAGCCTG	GTACCGTAAC	AATCCCTCAC	CCGTTCCAGG	1440
TCGGGGATCA	AGTGCTTGTC	AGACCCATC	GACCCAGCAG	CCTTGAGCCT	CGGTGGGAAA	1500
GCCCATACTT	GGTGTGCTG	ACTACCCCGA	CCGCGGTAAA	AGTCGATGGT	ATTGCTGCCT	1560
GGGTCCATGC	TTCTCACCTC	AAACCTGCAC	CACCTTCGGC	ACCAGATGAG	TCCTGGGAGC	1620
TGGAAAAGAC	TGATCATCCT	CTTAAGCTGC	GTATTCCGGC	GCGGCGGGAC	GAGTCTGCAA	1680
AATAAGAAC	CCCACAGCC	CATGACCCTC	ACTTGGCAGG	TACTGTCCCA	AACTGGAGAC	1740
GTTGTCTGGG	ATACAAAGGC	AGTCCAGCCC	CCTTGGACTT	GGTGGCCAC	ACTTAAACCT	1800
GATGTATGTG	CCTTGGCGGC	TAGTCTTGAG	TCCTGGGATA	TCCCGGGAAC	CGATGTCTCG	1860
TCCTCTAAAC	GAGTCAGACC	TCCGGACTCA	GACTATACTG	CCGCTTATAA	GCAAAATCAC	1920
TGGGGAGCCA	TAGGGTGCAG	CTACCCTCGG	GCTAGGACTA	GAATGGCAAG	CTCTACCTTC	1980
TACGTATGTC	CCCGGGATGG	CCGGACCTTT	TCCAGAGCTA	GAAGGTGCGG	GGGGTCTGAA	2040
TCCCTATACT	GTAAAGAATG	GGATTGTGAG	ACCACGGGGA	CCGGTTATTG	GCTATCTAAA	2100
TCCTCAAAAG	ACCTCATAAC	TGTAAATGG	GACCAAAATA	GCGAATGGAC	TCAAAATTTT	2160
CAACAGTGTC	ACCAAGCCGG	CTGGTGTAAC	CCCCTTAAAA	TAGATTTTAC	AGACAAAGGA	2220
AAATTATCCA	AGGACTGGAT	AACGGGAAAA	ACCTGGGGAT	TAAGATTCTA	TGTGTCTGGA	2280
CATCCAGGCG	TACAGTTCAC	CATTGCTTAA	AAAATCACCA	ACATGCCAGC	TGTGGCAGTA	2340
GGTCTTGACC	TCGTCTTGT	GGAACAAGGA	CCTCCTAGAA	CGTCCCTCGC	TCCTCCACCT	2400
CCTCTTCCCC	CAAGGAAGC	GCCACCGCCA	TCTCTCCCGG	ACTCTAACTC	CACAGCCCTG	2460
GCGACTAGTG	CACAACTCC	CACGGTGAGA	AAAACAATTG	TTACCCATAA	CACTCCGCTT	2520
CCCACCACAG	GCGACAGACT	TTTTGATCTT	GTGACGGGGG	CCTTCCTAAC	CTTAAATGCT	2580
ACCAACCCAG	GGGCCACTGA	GTCTTGCTGG	CTTTGTTTGG	CCATGGGCCC	CCCTTATTAT	2640
GAAGCAATAG	CCTCATCAGG	AGAGGTGCGC	TACTCCACCG	ACCTTGACCG	GTGCCGCTGG	2700
GGGACCCAAG	GAAAGCTCAC	CCTCACTGAG	GTCTCAGGAC	ACGGGTGTG	CATAGGAAGA	2760
GTGCCCTTTA	CCCATCAGCA	TCTCTGCAAT	CAGACCCTAT	CCATCAATTC	CTCCGGAGAC	2820
CATCAGTATC	TGCTCCCCTC	CAACCATAGC	TGGTGGGCTT	GCAGCACTGG	CCTCACCCCT	2880
TGCCCTCTCCA	CCTCAGTTT	TAATCAGACT	AGAGATTTCT	GTATCCAGGT	CCAGCTGATT	2940
CCTCGCATCT	ATTACTATCC	TGAAGAAGTT	TTGTTACAGG	CCTATGACAA	TTCTCACCCC	3000
AGGACTAAAA	GAGAGGCTGT	CTCACTTACC	CTAGCTGTTT	TACTGGGGTT	GGGAATCACG	3060
GCGGGAATAG	GTAAGTGGTT	AATGCGCTTA	ATTAAAGGAC	CTATAGACCT	CCAGCAAGGC	3120
CTGACAAGCC	TCCAGATCGC	CATAGATGCT	GACCTCCGGG	CCCTCCAAGA	CTCAGTCAGC	3180
AAGTTAGAGG	ACTCACTGAC	TTCCCTGTCC	GAGGTAGTGC	TCCAAAATAG	GAGAGGCCCT	3240
GACTTGCTGT	TTCTAAAAGA	AGGTGGCCTC	TGTGCGGCCC	TAAAGGAAGA	GTGCTGTTTT	3300
TACATAGACC	ACTCAGGTGC	AGTACGGGAC	TCCATGAAAA	AACTCAAAGA	AAAAGTGGAT	3360
AAAAGACAGT	TAGAGCGCCA	GAAAAGCCAA	AACCTGGTATG	AAGGATGGTT	CAATAACTCC	3420
CCTTGGTTCA	CTACCCTGCT	ATCAACCATC	GCTGGGCCCC	TATTACTCCT	CCTTCTGTGT	3480
CTCATCTCG	GGCATGCAT	CATCAATCGA	TTAGTTCAAT	TTGTTAAAGA	CAGGATCTCA	3540
GTAGTCCAGG	CTTTAGTCTT	GACTCAACAA	TACCACCAGC	TAAAGCCTAT	AGAGTACGAG	3600
CCATAGGGCG	CCTAGTGTG	ACAATTAATC	ATCGGCATAG	TATACGGCAT	AGTATAATAC	3660
GACTCACTAT	AGGAGGGCCA	CCATGGCCAA	GTTGACCAAT	GCCGTTCCGG	TGCTCACCCG	3720
GCCGACGTC	GCCGGAGCGG	TCGAGTTCTG	GACCGACCGG	CTCGGGTTCT	CCCGGACTTT	3780
CGTGGAGGAC	GACTTCGCGG	GTGTGGTCCG	GGACGACGTG	ACCCTGTTCA	TCAGCGCGGT	3840
CCAGGACCCAG	GTGGTCCCGG	ACAACACCTT	GGCCTGGGTG	TGGGTGCGCG	GCCTGGACGA	3900
GCTGTACGCC	GAGTGGTCCG	AGGTGCTGTC	CACGAATTTT	CGGGACGCCT	CCGGGCCGGC	3960
CATGACCCAG	ATCGGCGAGC	AGCCGTGGGG	GCGGGAGTTC	GCCCTGCGCG	ACCCGGCCCG	4020
CAACTGCGTG	CACCTCGTGG	CCGAGGAGCA	GGACTGANNN	NCGGACCGGT	CGACTTGTTA	4080

Figure 11. FBdelPGASAF Sequence

2

ACTTGTTTAT	TGCAGCTTAT	AATGGTTACA	AATAAAGCAA	TAGCATCACA	AATTTACAA	4140
ATAAAGCATT	TTTTTCACTG	CATTCTAGTT	GTGGTTTGTG	CAAACCTCATC	AATGTATCTT	4200
ATCATGTCTG	GATCCAGATC	TGGGCCCCATG	CGGCCGCGGA	TCGATNNNNA	CATGTAGCA	4260
AAAGGCCAGC	AAAAGGCCAG	GAACCGTAAA	AAGGCCGCGT	TGCTGGCGTT	TTTCCATAGG	4320
CTCCGCCCCC	CTGACGAGCA	TCACAAAAAT	CGACGCTCAA	GTCAGAGGTG	GCGAAACCCG	4380
ACAGGACTAT	AAAGATACCA	GGCGTTTCCC	CCTGGAAGCT	CCCTCGTGCG	CTCTCCTGTT	4440
CCGACCCTGC	CGCTTACCGG	ATACCTGTCC	GCCTTTCTCC	CTTCGGGAAG	CGTGGCGCTT	4500
TCTCAATGCT	CACGCTGTAG	GTATCTCAGT	TCGGTGTAGG	TCGTTGCGTC	CAAGCTGGGC	4560
TGTGTGCACG	AACCCCCCGT	TCAGCCCGAC	CGCTGCGCCT	TATCCGGTAA	CTATCGTCTT	4620
GAGTCCAACC	CGGTAAGACA	CGACTTATCG	CCACTGGCAG	CAGCCACTGG	TAACAGGATT	4680
AGCAGAGCGA	GGTATGTAGG	CGGTGCTACA	GAGTTCTTGA	AGTGGTGGCC	TAACACGGC	4740
TACACTAGAA	GGACAGTATT	TGGTATCTGC	GCTCTGCTGA	AGCCAGTTAC	CTTCGGAAAA	4800
AGAGTTGGTA	GCTCTTGATC	CGGCAACAA	ACCACCGCTG	GTAGCGGTGG	TTTTTTTGTT	4860
TGCAAGCAGC	AGATTACGCG	CAGAAAAAAA	GGATCTCAAG	AAGATCCTTT	GATCTTTTCT	4920
ACGGGGTCTG	ACGCTCAGTG	GAACGAAAAC	TCACGTTAAG	GGATTTTGGT	CATGAGATTA	4980
TCAAAAAGGA	TCTTCACCTA	GATCCTTTTA	AATTAAAAAT	GAAGTTTAA	ATCAATCTAA	5040
AGTATATATG	AGTAAACTTG	GTCTGACAGT	TACCAATGCT	TAATCAGTGA	GGCACCCTATC	5100
TCAGCGATCT	GTCTATTTTCG	TTCATCCATA	GTTGCCTGAC	TCCCCGTCGT	GTAGATAACT	5160
ACGATACGGG	AGGGCTTACC	ATCTGGCCCC	AGTGCTGCAA	TGATACCGCG	AGACCCACGC	5220
TCACCGGCTC	CAGATTTATC	AGCAATAAAC	CAGCCAGCCG	GAAGGGCCGA	GCGCAGAAGT	5280
GGTCCTGCAA	CTTTATCCGC	CTCCATCCAG	TCTATTAATT	GTTGCCGGGA	AGCTAGAGTA	5340
AGTAGTTTCG	CAGTTAATAG	TTTGCCCAAC	GTTGTTGCCA	TTGCTACAGG	CATCGTGGTG	5400
TCACGCTCGT	CGTTTGGTAT	GGCTTCATTC	AGCTCCGGTT	CCCAACGATC	AAGGCGAGTT	5460
ACATGATCCC	CCATGTTGTG	CAAAAAAGCG	GTTAGCTCCT	TCGGTCCTCC	GATCGTTGTC	5520
AGAAGTAAGT	TGGCCGCACT	GTTATCACTC	ATGGTTATGG	CAGCACTGCA	TAATTCTCTT	5580
ACTGTCTATG	CATCCGTAAG	ATGCTTTTCT	GTGACTGGTG	AGTACTCAAC	CAAGTCATTC	5640
TGAGAATAGT	GTATGCGGCG	ACCGAGTTGC	TCTTGCCCCG	CGTCAATACG	GGATAATACC	5700
GCGCCACATA	GCAGAACTTT	AAAAGTGCTC	ATCATTTGGAA	AACGTTCTTC	GGGGCGAAAA	5760
CTCTCAAGGA	TCTTACCGCT	GTTGAGATCC	AGTTGATGAT	AACCCACTCG	TGCACCCAAC	5820
TGATCTTCAG	CATCTTTTAC	TTTCACCAGC	GTTTCTGGGT	GAGCAAAAAC	AGGAAGGCAA	5880
AATGCCGCAA	AAAAGGGAAT	AAGGGCGACA	CGGAAATGTT	GAATACTCAT	ACTCTTCTTT	5940
TTTCAATATT	ATTGAAGCAT	TTATCAGGGT	TATTGTCTCA	TGAGCGGATA	CATATTTGAA	6000
TGTATTTAGA	AAAATAAACA	AATAGGGGTT	CCGCGCACAT	TTCCCCGAAA	AGTGCCACCT	6060
GACGTCTAAG	AAACCATTAT	TATCATGACA	TTAACCTATA	AAAATAGGCG	TATCAGGAGG	6120
CCCTTTCTGTC	TCGCGCGTTT	CGGTGATGAC	GGTGAAAACC	TCTGACACAT	GCAGCTCCCG	6180
GAGACGGTCA	CAGCTTGTCT	GTAAGCGGAT	GCCGGGAGCA	GACAAGCCCG	TCAGGGCGCG	6240
TCAGCGGGTG	TTGGCGGGTG	TCGGGGCTGG	CTTAACCTATG	CGGCATCAGA	GCAGATTGTA	6300
CTGAGAGTGC	AC					6312

Figure 12. FBdelPRDSA F Sequence

1

CATATGCGGT	GTGAAATACC	GCACAGATGC	GTAAGGAGAA	AATACCGCAT	CAGGCGCCAT	60
TCGCCATTCA	GGCTGCGCAA	CTGTTGGGAA	GGGCGATCGG	TGCGGGCCTC	TTGCTATTA	120
CGCCAGCTGG	CGAAAGGGGG	ATGTGCTGCA	AGGCGATTAA	GTTGGGTAAC	GCCAGGGTTT	180
TCCCAGTCAC	GACGTTGTAA	AACGACGGCC	AGTGAATTCC	GATTAGTTCA	ATTGTTTAAA	240
GACAGGATCT	CAGTAGTCCA	GGCTTTAGTC	CTGACTCAAC	AATACCACCA	GCTAAAAACA	300
CTAGAATACG	AGCCACAATA	AATAAAAGAT	TTTATTTAGT	TTCCAGAAAA	AGGGGGGAAT	360
GAAAGACCCC	ACCAAATTGC	TTAGCCTGAT	AGCCGCAGTA	ACGCCATTTT	GCAAGGCATG	420
GAAAAATACC	AAACCAAGAA	TAGAGAAGTT	CAGATCAAGG	GCGGGTACAC	GAAAACAGCT	480
AACGTTGGGC	CAACACAGAT	ATCTGCGGTG	AGCAGTTTCG	GCCCCGGCCC	GGGGCCAAGA	540
ACAGATGGTC	ACCGCGGTTT	GGCCCCGGCC	CGGGGCCAAG	AACAGATGGT	CCCCAGATAT	600
GGCCCAACCC	TCAGCAGTTT	CTTAAGACCC	ATCAGATGTT	TCCAGGCTCC	CCCAAGGACC	660
TGAAATGACC	CTGTGCCTTA	TTTGAATTAA	CCAATCAGCC	TGCTTCTCGC	TTCTGTTTCG	720
GCGCTTCTGC	TTCCCGAGCT	CTATAAAAGA	GCTCACAACC	CCTCACTCGG	CGCGCCAGTC	780
CTCCGATAGA	CTGAGTCGCC	CGGGTACCCG	TGTATCCAAT	AAATCCTCTT	GCTGTTGCAT	840
CCGACTCGTG	GTCTCGCTGT	TCCTTGGGAG	GGTCTCCTCA	GAGTGATTGA	CTACCCGTCT	900
CGGGGGTCTT	TCATTTGGGG	GCTCGTCCGG	GATCTGGAGA	CCCCTGCCCA	GGGACCACCG	960
ACCCACCACC	GGGAGGTAAG	CTGGCCAAGA	TCCCCCGGGC	TGCAGGAATT	TATGAAATCC	1020
TTTATGGGGG	ACCCCCCCTT	TTGTCAACCT	TGCTCAATTC	CTTCTCCCCC	TCCGATCCTA	1080
AGACTGATTT	ACAAGCCCCG	CTAAAAGGGC	TGCAAGGCGT	GCAGGCCCAA	ATCTGGACAC	1140
CCCTGGCCGA	ATTGTACCGG	CCAGGACATC	CACAAACTAG	CCACCCATTT	CAGGTGGGAG	1200
ACTCCGTGTA	GCTCCGGCGG	CACCGCTCTC	AAGGATTGGA	GCCTCGTTGG	AAGGGACCTT	1260
ACATCGTCCT	GCTGACCACG	CCCACCGCCA	TAAAGGTTGA	CGGGATCGCC	GCCTGGATTTC	1320
ACGCATCGCA	CGCCAAGGCA	GCCCCAAAAA	CCCCTGGACC	AGAAACTCCC	AAAACCTGGA	1380
AGCTCCGCCG	TTGGGAGAAC	CCTCTTAAGA	TAAGACTCTC	CCGTGTCTGA	CTGCTAATCC	1440
ACCTTGTCCC	TGTACTAACC	CAAAATGAAA	CTCCCAACAG	GAATGGTCAT	TTTATGTAGC	1500
CTAATAATAG	TTCCGGGCAGG	GTTTGACGAC	CCCCGCAAGG	CTATCGCATT	AGTACAAAAA	1560
CAACATGGTA	AACCATGCGA	ATGCAGCGGA	GGGCAGGTAT	CCGAGGCCCC	ACCGAATCC	1620
ATCCAACAGG	TAACCTGCCC	AGGCAAGACG	GCCTACTTAA	TGACCAACCA	AAAATGGAAA	1680
TGCAGAGTCA	CTCCAAAAT	CTCACCTAGC	GGGGGAGAAC	TCCAGAACTG	CCCCTGTAAC	1740
ACTTTCCAGG	ACTCGATGCA	CAGTCTTGTG	TATACTGAAT	ACCGGCAATG	CAGGCGAATT	1800
AATAAGACAT	ACTACACGGC	CACCTTGCTT	AAAATACGGT	CTGGGAGCCT	CAACGAGGTA	1860
CAGATATTAC	AAAACCCCAA	TCAGTCCCTA	CAGTCCCCTT	GTAGGGGCTC	TATAAATCAG	1920
CCCGTTTGCT	GGAGTGGCAC	AGCCCCATC	CATATCTCCG	ATGGTGGAGG	ACCCCTCGAT	1980
ACTAAGAGAG	TGTGGACAGT	CCAAAAAAGG	CTAGAACAAA	TTCATAAGGC	TATGACTCCT	2040
GAACCTCAAT	ACCACCCCTT	AGCCCTGCCC	AAAGTCAGAG	ATGACCTTAG	CCTTGATGCA	2100
CGGACTTTTG	ATATCCTGAA	TACCACTTTT	AGGTTACTCC	AGATGTCCAA	TTTTAGCCTT	2160
GCCCAAGATT	GTTGGCTCTG	TTTAAAACTA	GGTACCCCTA	CCCCTCTTGC	GATACCCAT	2220
CCCTCTTTAA	CCTACTCCCT	AGCAGACTCC	CTAGCGAATG	CCTCCTGTCA	GATTATACCT	2280
CCCCTCTTGG	TTCAACCGAT	GCAGTTCCTC	AACTCGTCCT	GTTTATCTTC	CCCTTTCATT	2340
AACGATACGG	AACAATAGA	CTTAGGTGCA	GTCACCTTTA	CTAACTGCAC	CTCTGTAGCC	2400
AATGTCAGTA	GTCCTTTATG	TGCCCTAAAC	GGGTCACTGT	TCCTCTGTGG	AAATAACATG	2460
GCATACACCT	ATTTACCCCA	AAACTGGACC	AGACTTTGCG	TCCAAGCCTC	CCTCCTCCCC	2520
GACATTGACA	TCAACCCGGG	GGATGAGCCA	GTCCCCATTC	CTGCCATTGA	TCATTATATA	2580
CATAGACCTA	AACGAGCTGT	ACAGTTCATC	CCTTTACTAG	CTGGACTGGG	AATCACCAGCA	2640
GCATTACCCA	CCGGAGCTAC	AGGCCTAGGT	GTCTCCGTCA	CCCAGTATAC	AAAATTATCC	2700
CATCAGTTAA	TATCTGATGT	CCAAGTCTTA	TCCGGTACCA	TACAAGATT	ACAAGACCAG	2760
GTAGACTCGT	TAGCTGAAGT	AGTTCTCCAA	AATAGGAGGG	GACTGGACCT	ACTAACGGCA	2820
GAACAAGGAG	GAATTTGTTT	AGCCTTACAA	GAAAAATGCT	GTTTTTTATGC	TAACAAGTCA	2880
GGAATTGTGA	GAAACAAAAT	AAGAACCCTA	CAAGAAGAAAT	TACAAAAACG	CAGGGAAGAGC	2940
CTGGCAACCA	ACCCTCTCTG	GACCGGGCTG	CAGGGCTTTC	TTCCGTACCT	CCTACCTCTC	3000
CTGGGACCCC	TACTCACCTT	CCTACTCATA	CTAACCATTG	GGCCATGCGT	TTTCAGTCCG	3060
CTCATGGCCT	TCATTAATGA	TAGACTTAAT	GTTGTACATG	CCATGGTGCT	GGCCAGCAAA	3120
TACCAAGCAC	TCAAAGCTGA	GGAAGAAGCT	CAGGATTGAG	GCGCCTAGTG	TTGACAATTA	3180
ATCATCGGCA	TAGTATACGG	CATAGTATAA	TACGACTCAC	TATAGGAGGG	CCACCATGGT	3240
CAAGTTGACC	AGTGCCGTTT	CGGTGCTCAC	CGCGCGCGAC	GTCGCCGGAG	CGGTGAGTTC	3300
CTGGACCGAC	CGGCTCGGGT	TCTCCCGGGA	CTTCGTGGAG	GACGACTTCG	CCGGTGTGGT	3360
CCGGGACGAC	GTGACCCCTG	TCATCAGCGC	GGTCCAGGAC	CAGGTGGTGC	CGGACAACAC	3420
CCTGGCCTGG	GTGTGGGTGC	GCGGCCTGGA	CGAGCTGTAC	GCCGAGTGGT	CGGAGGTCGT	3480
GTCCACGAAC	TTCCGGGACG	CCTCCGGGGC	GGCCATGACC	GAGATCGGCG	AGCAGCCGTG	3540
GGGGCGGGAG	TTCGCCCTGC	GCGACCCGGC	CGGCAACTGC	GTGCACTTCG	TGGCCGAGGA	3600
GCAGGACTGA	NNNNCGGACC	GGTCGACTTG	TTAATCTGTT	TATTGCAGCT	TATAATGGTT	3660
ACAAATAAAG	CAATAGCATC	ACAAATTTCA	CAAATAAAGC	ATTTTTCCTA	CTGCATTCTA	3720
GTTGTGGTTT	GTCCAAACTC	ATCAATGTAT	CTTATCATGT	CTGGATCCAG	ATCTGGGCCC	3780
ATGCGGCGCG	GGATCGATNN	NNACATGTGA	GCAAAAGGCC	AGCAAAAGGC	CAGGAACCGT	3840
AAAAAGGCCG	CGTTGCTGGC	GTTTTTCCAT	AGGCTCCGCC	CCCCTGACGA	GCATCACAAA	3900
AATCGACGCT	CAAGTCAGAG	GTGGCGAAAC	CCGACGAGAC	TATAAAGATA	CCAGCGCTTT	3960
CCCCCTGGAA	GCTCCCTCGT	GCGCTCTCCT	GTTCCGACCC	TGCCGCTTAC	CGGATACCTG	4020
TCCGCTTTTC	TCCCTTCGGG	AAGCGTGGCG	CTTCTCAAT	GCTCAGCTG	TAGGTATCTC	4080

Figure 12. FBdelPRDSAF Sequence

2

AGTTCGGTGT	AGGTCGTTCTG	CTCCAAGCTG	GGCTGTGTGC	ACGAACCCCC	CGTTCAGCCC	4140
GACCGCTGCG	CCTTATCCGG	TAACTATCGT	CTTGAGTCCA	ACCCGGTAAG	ACACGACTTA	4200
TCGCCACTGG	CAGCAGCCAC	TGGTAACAGG	ATTAGCAGAG	CGAGGTATGT	AGGCGGTGCT	4260
ACAGAGTTCT	TGAAGTGGTG	GCCTAACTAC	GGCTACACTA	GAAGGACAGT	ATTTGGTATC	4320
TGCGCTCTGC	TGAAGCCAGT	TACCTTCGGA	AAAAGAGTTG	GTAGCTCTTG	ATCCGGCAAA	4380
CAAACCACCG	CTGGTAGCGG	TGGTTTTTTT	GTTTGCAAGC	AGCAGATTAC	GCGCAGAAAA	4440
AAAGGATCTC	AAGAAGATCC	TTTGATCTTT	TCTACGGGGT	CTGACGCTCA	GTGGAACGAA	4500
AACTCACGTT	AAGGGATTTT	GGTCATGAGA	TTATCAAAAA	GGATCTTCAC	CTAGATCCTT	4560
TTAAATTAAA	AATGAAGTTT	TAAATCAATC	TAAAGATAT	ATGAGTAAAC	TTGGTCTGAC	4620
AGTTACCAAT	GCTTAATCAG	TGAGGCACCT	ATCTCAGCGA	TCTGTCTATT	TCGTTTCATCC	4680
ATAGTTGCCT	GACTCCCCGT	CGTGTAGATA	ACTACGATAC	GGGAGGGCTT	ACCATCTGGC	4740
CCCAGTGCTG	CAATGATACC	GCGAGACCCA	CGCTCACCGG	CTCCAGATTT	ATCAGCAATA	4800
AACCAGCCAG	CCGGAAGGGC	CGAGCGCAGA	AGTGGTCTCTG	CAACTTTATC	CGCCTCCATC	4860
CAGTCTATTA	ATTGTTGCCG	GGAAGCTAGA	GTAAGTAGTT	CGCCAGTTAA	TAGTTTGCGC	4920
AACGTTGTTG	CCATTGCTAC	AGGCATCGTG	GTGTCACGCT	CGTCGTTTGG	TATGGCTTCA	4980
TTCAGCTCCG	GTTCCCAACG	ATCAAGGCGA	GTTACATGAT	CCCCCATGTT	GTGCAAAAAA	5040
GCGGTTAGCT	CCTTCGGTCC	TCCGATCGTT	GTCAGAAGTA	AGTTGGCCGC	AGTGTATATCA	5100
CTCATGGTTA	TGGCAGCACT	GCATAATTCT	CTTACTGTCA	TGCCATCCGT	AAGATGCTTT	5160
TCTGTGACTG	GTGAGTACTC	AACCAAGTCA	TTCTGAGAAT	AGTGTATGCG	GCGACCGAGT	5220
TGCTCTTGCC	CGGCGTCAAT	ACGGGATAAT	ACCGCGCCAC	ATAGCAGAAC	TTTAAAAGTG	5280
CTCATCATTG	GAAAACGTTT	TTCGGGGCGA	AAACTCTCAA	GGATCTTACC	GCTGTTGAGA	5340
TCCAGTTCGA	TGTAACCCAC	TCGTGCACCC	AACTGATCTT	CAGCATCTTT	TACTTTTACC	5400
AGCGTTTCTG	GGTGAGCAAA	AACAGGAAGG	CAAAATGCCG	CAAAAAAGGG	AATAAGGGCG	5460
ACACGGAAAT	GTTGAATACT	CATACTCTTC	CTTTTTCAAT	ATTATTGAAG	CATTTATCAG	5520
GGTTATTGTC	TCATGAGCGG	ATACATATTT	GAATGTATTT	AGAAAAATAA	ACAAATAGGG	5580
GTTCCGCGCA	CATTTCCCGG	AAAAGTGCCA	CCTGACGTCT	AAGAAACCAT	TATTATCATG	5640
ACATTAACCT	ATAAAAAATAG	GCGTATCACG	AGGCCCTTTC	GTCTCGCGCG	TTTCGGTGAT	5700
GACGGTGAAA	ACCTCTGACA	CATGCAGCTC	CCGGAGACGG	TCACAGCTTG	TCTGTAAGCG	5760
GATGCCGGGA	GCAGACAAGC	CCGTACGGGC	GCGTCAGCGG	GTGTTGGCGG	GTGTCGGGGC	5820
TGGCTTAACT	ATGCGGCATC	AGAGCAGATT	GTACTGAGAG	TGCAC		5865



Figure 13. hCMV10A1 Sequence

1

AGATCTCCCG	ATCCCCCTATG	GTCGACTCTC	AGTACAATCT	GCTCTGATGC	CGCATAGTTA	60
AGCCAGTATC	TGCTCCCTGC	TTGTGTGTTG	GAGGTGCGTG	AGTAGTGC GC	GAGCAAAATT	120
TAAGCTACAA	CAAGGCAAGG	CTTGACCGAC	AATTGCATGA	AGAATCTGCT	TAGGGTTAGG	180
CGTTTTGCGC	TGCTTCGCGA	TGTACGGGCC	AGATATACGC	GTTGACATTG	ATTATTGACT	240
AGTTATTAAT	AGTAATCAAT	TACGGGGTCA	TTAGTTCATA	GCCCATATAT	GGAGTTCCGC	300
GTTACATAAC	TTACGGTAAA	TGGCCCGCCT	GGCTGACCGC	CCAACGACCC	CCGCCCATTG	360
ACGTCAATAA	TGACGTATGT	TCCCATAGTA	ACGCCAATAG	GGACTTTCCA	TTGACGTCAA	420
TGGGTGGACT	ATTTACGGTA	AACTGCCCAC	TTGGCAGTAC	ATCAAGTGT	TCATATGCCA	480
AGTACGCCCC	CTATTGACGT	CAATGACGGT	AAATGGCCCG	CCTGGCATT	TGCCGAGTAC	540
ATGACCTTAT	GGGACTTTCC	TACTTGGCAG	TACATCTACG	TATTAGTCAT	CGCTATTACC	600
ATGGTGATGC	GGTTTTGGCA	GTACATCAAT	GGGCGTGGAT	AGCGGTTTGA	CTCACGGGGA	660
TTTCCAAGT	TCCACCCCAT	TGACGTCAAT	GGGAGTTTGT	TTTGGCACCA	AAATCAACGG	720
GACTTTCCAA	AATGTCGTAA	CAACTCCGCC	CCATTGACGC	AAATGGGCGG	TAGGCGTGTA	780
CGGTGGGAGG	TCTATATAAG	CAGAGCTCTC	TGGCTAACTA	GAGAACCCAC	TGCTTAACTG	840
GCTTATCGAA	ATGTCGACTG	AGAACCTTCA	GGTGAGTTTG	GGGACCCTTG	ATTGTTCTTT	900
CTTTTTCGCT	ATTGTAAAT	TCATGTTATA	TGGAGGGGGC	AAAGTTTTC	GGGTGTTGTT	960
TAGAATGGGA	AGATGTCCTT	TGTATACCA	TGGACCCTCA	TGATAATTTT	GTTTCTTTCA	1020
CTTTCTACTC	TGTTGACAAC	CATTGTCTCC	TCTTATTTTC	TTTTCAATTT	CTGTAATTTT	1080
TTCGTTAAAC	TTTAGCTTGC	ATTTGTAACG	AATTTTTTAA	TTCACCTTTG	TTTATTTGTC	1140
AGATTGTAAG	TACTTTCTCT	AATCACTTTT	TTTTCAAGGC	AATCAGGGTA	TATTATATTG	1200
TACTTTCAGCA	CAGTTTTCAG	GAACAATTGT	TATAATTAAA	TGATAAGGTA	GAATATTTCT	1260
GCATATAAAT	TCTGGCTGGC	GTGGAAATAT	TCTTATTTGT	AGAAACAAC	ACATCCTGGT	1320
CATCATCCTG	CCTTTCTCTT	TATGGTTACA	ATGATATACA	CTGTTTGAGA	TGAGGATAAA	1380
ATACTCTGAG	TCCAAACCGG	GCCCCCTCTG	TAACCATGTT	CATGCCTTCT	TCTTTTTTCT	1440
ACAGCTCCCTG	GGCAACGTGC	TGGTTGTTGT	GCTGTCTCAT	CATTTTGGCA	AGGATCGGCC	1500
GGAACAGCAT	CAGGACCGAC	ATGGAAGGTC	CAGCGTTCTC	AAAACCCCTT	AAAGATAAGA	1560
TTAACCCGTG	GAAGTCCTTA	ATGGTCATGG	GGGTCTATTT	AAGAGTAGGG	ATGGCAGAGA	1620
GCCCCCATCA	GGTCTTTAAT	GTAACCTGGA	GAGTCACCAA	CCTGATGACT	GGGCGTACCG	1680
CCAATGCCAC	CTCCCTTTTA	GGAACGTGAC	AAGATGCCTT	CCCAAGATTA	TATTTTGATC	1740
TATGTGATCT	GGTCGGAGAA	GAGTGGGACC	CTTCAGACCA	GGAACCATAT	GTCGGGTATG	1800
GCTGCAATA	CCCCGGAGGG	AGAAAGCGGA	CCCGGACTTT	TGACTTTTAC	GTGTGCCCTG	1860
GGCATACCGT	AAAATCGGGG	TGTGGGGGGC	CAAGAGAGGG	CTACTGTGGT	GAATGGGGTT	1920
GTGAAACAC	CGGACAGGCT	TACTTGGAAGC	CCACATCATC	ATGGGACCTA	ATCTCCCTTA	1980
AGCGCGGTAA	CACCCCTTGG	GACACGGGAT	GGCTTGTGGC	CCCTGCTACG	CCCTGCTACG	2040
ACCTCTCCAA	AGTATCCAAT	TCCTTCCAAG	GGGCTACTCG	AGGGGGCAGA	TGCAACCCCTC	2100
TAGTCTTAGA	ATTCACTGAT	GCAGGAAAAA	AGGCTAATTG	GGACGGGCCC	AAATCGTGGG	2160
GACTGAGACT	GTACCGGACA	GGAACAGATC	CTATTACCAT	GTTCTCCCTG	ACCCGCCAGG	2220
TCCTCAATAT	AGGGCCCCGC	ATCCCCATTG	GGCCTAATCC	CGTGATCACT	GGTCACTAC	2280
CCCCCTCCCG	ACCCGTGCAG	ATCAGGCTCC	CCAGGCTTCC	TCAGCCTCCT	CCTACAGGCG	2340
CAGCCTCTAT	AGTCCCTGAG	ACTGCCCCAC	CTTCTCAACA	ACCTGGGACG	GGAGACAGGC	2400
TGCTAAACCT	GGTGAAGGA	GCCTATCAGG	CGCTTAACTT	CACCAATCCC	GACAAGACCC	2460
AAGAATGTTG	GCTGTGCTTA	GTGTGCGGAC	CTCCTTATTA	CGAAGGAGTA	GCGGTGCTGG	2520
GCACTTATAC	CAATCATTTCT	ACCGCCCCGG	CCAGCTGTAC	GGCCACTTCC	CAACATAAGC	2580
TTACCCATATC	TGAAGTGACA	GGACAGGGCC	TATGCATGGG	AGCACTACCT	AAAACCTACC	2640
AGGCCCTTATG	TAACACCACC	CAAAGTGCCG	GCTCAGGATC	CTACTACCTT	GCAGACCCCG	2700
CTGGAACAAT	GTGGGCTTGT	AGCACTGGAT	TGACTCCCTG	CTTGTCCACC	ACGATGCTCA	2760
ATCTAACCAC	AGACTATTGT	GTATTAGTTG	AGCTCTGGCC	CAGAATAATT	TACCACTCCC	2820
CCGATTATAT	GTATGGTCAG	CTTGAACAGC	GTACCAAATA	TAAGAGGGAG	CCAGTATCGT	2880
TGACCCCTGGC	CCTTCTGCTA	GGAGGATTAA	CCATGGGAGG	GATTGCAGCT	GGAATAGGGA	2940
CGGGGACCAC	TGCCCTAATC	AAAACCCAGC	AGTTTGAGCA	GCTTCACGCC	GCTATCCAGA	3000
CAGACCTCAA	CGAAGTCGAA	AAATCAATTA	CCAACCTAGA	AAAGTCACTG	ACCTCGTTGT	3060
CTGAAGTAGT	CCTACAGAAC	CGAAGAGGCC	TAGATTTGCT	CTTCCTAAAA	GAGGGAGGTC	3120
TCTGCGCAGC	CCTAAAAGAA	GAATGTTGTT	TTTATGCAGA	CCACACGGGA	CTAGTGAGAG	3180
ACAGCATGGC	CAAACCTAAGG	GAAAGGCTTA	ATCAGAGACA	AAAACCTATT	GAGTCAGGCC	3240
AAGGTTGGTT	CGAAGGGCAG	TTTAATAGAT	CCCCCTGGTT	TACCACCTTA	ATCTCCACCA	3300
TCATGGGACC	TCTAATAGTA	CTCTTACTGA	TCTTACTCTT	TGGACCCTGC	ATTCTCAATC	3360
GATTAGTTCA	ATTTGTTAAA	GACAGGATCT	CAGTAGTCCA	GGCTTTAGTC	CTGACTCAAC	3420
AATACCACCA	GCTAAAGCCT	ATAGATACG	AGCCATAGGG	CGCCTAGTGT	TGACAATAAA	3480
TCATCGGCAT	AGTATACGGC	ATAGTATAAT	ACGACTCACT	ATAGGAGGGC	CACCATGGCC	3540
AAGTTGACCA	GTGCCGTTC	GGTGCTCACC	GCGCGCGACG	TCGCCGGAGC	GGTCGAGTTC	3600
TGGACCGACC	GGCTCGGGTT	CTCCCGGGAC	TTCGTGGAGG	ACGACTTCGC	CGGTGTGGTC	3660
CGGGACGACG	TGACCTGTT	CATCAGCGCG	GTCAGGACCC	AGGTGGTGCC	GGACAACACC	3720
CTGGCCTGGG	TGTGGGTGCG	CGGCCTGGAC	GAGCTGTACG	CCGAGTGGTC	GGAGGTGCTG	3780
TCCACGAAC	TCCGGGACGC	CTCCGGGCGG	GCCATGACCG	AGATCGGCGA	GCAGCCGTGG	3840
GGGCGGGAGT	TGCCCTTGCG	CGACCCGGCC	GGCAACTGCG	TGCACTTCGT	GGCCGAGGAG	3900
CAGGACTGAN	NNNCGGACCG	GTCGA				3925

## INTERNATIONAL SEARCH REPORT

International Application No

PCT/GB 96/02061

## A. CLASSIFICATION OF SUBJECT MATTER

IPC 6 C12N15/86 C12N5/10 C12N15/67

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 6 C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JOURNAL OF VIROLOGY 69 (7). 1995. 4086-4094. ISSN: 0022-538X, July 1995, XP002023654 LUUKKONEN B G M ET AL: "Efficiency of reinitiation of translation on human immunodeficiency virus type 1 mRNAs is determined by the length of the upstream open reading frame and by intercistronic distance." see the whole document ---	1-29
A	VIROLOGY (1995), 208(1), 215-25 CODEN: VIRLAX; ISSN: 0042-6822, 1 April 1995, XP002023655 HERZOG, ETIENNE ET AL: "Translation of the second gene of peanut clump virus RNA 2 occurs by leaky scanning in vitro" see the whole document ---	1-29
-/--		

☒ Further documents are listed in the continuation of box C.☒ Patent family members are listed in annex.

## \* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&amp;" document member of the same patent family

Date of the actual completion of the international search

23 January 1997

Date of mailing of the international search report

12.02.97

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax (+31-70) 340-3016

Authorized officer

Hornig, H

## INTERNATIONAL SEARCH REPORT

International Application No

PCT/GB 96/02061

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	J. VIROL. (1993), 67(8), 4886-95 CODEN: JOVIAM;ISSN: 0022-538X, August 1993, XP000616337 FOUILLOT, NATHALIE ET AL: "Translation of the hepatitis B virus P gene by ribosomal scanning as an alternative to internal initiation" see the whole document ---	1-29
A	VIROLOGY, vol. 188, no. 1, May 1992, ACADEMIC PRESS, INC.,NEW YORK, US, pages 342-352, XP002023656 C.-G. LIN AND S.J. LO: "Evidence for involvement of a ribosomal leaky scanning mechanism in the translation of the hepatitis B virus Pol gene from the viral pregenome RNA" see the whole document ---	1-29
A	VIROLOGY, vol. 185, no. 2, December 1991, ACADEMIC PRESS, INC.,NEW YORK, US, pages 862-866, XP000616129 F.-L. COSSET ET AL.: "Newcastle disease virus (NDV) vaccine based on immunization with avian cells expressing the NDV hemagglutinin-neuraminidase glycoprotein" cited in the application see the whole document ---	1-29
A	MOL. CELL. BIOL., vol. 7, no. 10, October 1987, ASM WASHINGTON, DC,US, pages 3438-3445, XP000616288 M. KOZAK: "Effects of intercistronic length on the efficiency of reinitiation by eucaryotic ribosomes" cited in the application see the whole document ---	1-29
A	EXPRIMENTAL CELL RESEARCH, vol. 197, no. 2, December 1991, ACADEMIC PRESS INC., NEW YORK, US, pages 229-233, XP000616323 M. IZUMI ET AL.: "Blasticidin S-resistance gene (bsr): a novel selectable marker for mammalian cells" cited in the application see the whole document ---	1-29

-/--

## INTERNATIONAL SEARCH REPORT

International Application No

PCT/GB 96/02061

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	NATURE (LONDON) (1984), 309(5963), 82-5 CODEN: NATUAS;ISSN: 0028-0836, 3 May 1984, XP002023657 LIU, CHUNG CHENG ET AL: "Initiation of translation at internal AUG codons in mammalian cells" see the whole document ---	1-29
A	WO,A,94 24870 (BIOTRANSPLANT INC ;GEN HOSPITAL CORP (US); LE GUERN CHRISTIAN A (U) 10 November 1994 see the whole document ---	1-29
A	WO,A,93 03143 (ANDERSON W FRENCH ;MORGAN RICHARD A (US); COUTURE LARRY (US)) 18 February 1993 see the whole document ---	1-29
A	WO,A,94 23048 (US HEALTH ;EIDEN MARYBETH V (US); WILSON CAROLYN A (US); DEACON NI) 13 October 1994 see the whole document ---	1-29
A	PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA, vol. 85, no. 17, 1 September 1988, pages 6460-6464, XP000569693 DANOS O ET AL: "SAFE AND EFFICIENT GENERATION OF RECOMBINANT RETROVIRUSES WITH AMPHOTROPIC AND ECOTROPIC HOST RANGES" see the whole document ---	1-29
P,X	J. VIROL. ( 1995 ), 69(12), 7430-6 CODEN: JOVIAM;ISSN: 0022-538X, December 1995, XP000569527 COSSET, FRANCOIS-LOIC ET AL: "High-titer packaging cells producing recombinant retroviruses resistant to human serum" see the whole document -----	1-29

# INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/GB 96/02061

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
W0-A-9424870	10-11-94	AU-A- 6230394 CA-A- 2162056 EP-A- 0706319	21-11-94 10-11-94 17-04-96
W0-A-9303143	18-02-93	CA-A- 2114416 EP-A- 0598029 JP-T- 6509713	18-02-93 25-05-94 02-11-94
W0-A-9423048	13-10-94	AU-A- 6903194 CA-A- 2160034 EP-A- 0699240 JP-T- 8511156	24-10-94 13-10-94 06-03-96 26-11-96